```
                    GenCore version 5.1.6
          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   July 29, 2004, 08:20:25 ; Search time 93 Seconds
                                     (without alignments)
                            2111.511 Million cell updates/sec

Title:        US-10-719-885-2
Perfect score: 3669
Sequence:     1 MKRGKDEEKILEPMFPRLHV.........AARIFQSIQEERKRYDSSKP 695

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:     1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters:    1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   A_Geneseq_29Jan04:*
          1: geneseqp1980s:*
          2: geneseqp1990s:*
          3: geneseqp2000s:*
          4: geneseqp2001s:*
          5: geneseqp2002s:*
          6: geneseqp2002as:*
          7: geneseqp2003bs:*
          8: geneseqp2004s:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3669	100.0	695	6	AAY45097	Aay45097 Arabidops
2	3327	90.7	695	6	AAO15964	Aao15964 Arabidops
3	3321	90.5	695	6	AAO15963	Aao15963 Arabidops
4	197	5.4	971	7	ADE56304	Ade56304 Human Pro
5	193.5	5.3	802	5	ABG93215	Abg93215 S. cerevi
6	193.5	5.3	802	6	ABR53105	Abr53105 Protein s
7	192	5.2	1503	4	ABB60944	Abb60944 Drosophil
8	189.5	5.2	738	5	ABG93140	Abg93140 S. cerevi
9	189.5	5.2	1130	4	ABP73921	Abp73921 Candida a
10	186.5	5.1	1138	5	ABC05434	Abc05434 Human bre
11	186	5.1	771	4	ABB58529	Abb58529 Drosophil
12	180	4.9	1594	4	ABB63688	Abb63688 Drosophil
13	179.5	4.9	1133	6	AAE36097	Aae36097 Human Mas
14	179.5	4.9	1133	5	ADC59316	Adc59316 Human pol
15	178	4.8	1420	4	ABB63410	Abb63410 Drosophil
16	177	4.8	748	5	ABB92874	Abb92874 Herbicida
17	176	4.8	2237	4	ABB64977	Abb64977 Drosophil
18	174.5	4.8	757	4	AAG85017	Aag85017 Shrimp wh
19	172	4.7	1187	4	ABB67666	Abb67666 Drosophil
20	172	4.7	3502	4	ABB58382	Abb58382 Drosophil
21	171.5	4.7	1596	6	AAE36095	Aae36095 Drosophil
22	170.5	4.6	843	4	AAM79967	Aam79967 Human pro
23	170	4.6	2328	4	ABB71657	Abb71657 Drosophil
24	168.5	4.6	333	4	ABB62536	Abb62536 Drosophil
25	168.5	4.6	506	4	ABB71899	Abb71899 Drosophil
26	168.5	4.6	1125	7	ADC37561	Adc37561 Human nuc
27	166.5	4.5	1441	7	ADC23487	Adc23487 Receptor
28	165.5	4.5	1366	4	ABB66921	Abb66921 Drosophil
29	165.5	4.5	1920	4	ABG64441	Abg64441 Drosophil
30	165	4.5	611	5	ABG93274	Abg93274 C. albica
31	165	4.5	1790	4	ABB71739	Abb71739 Drosophil
32	164.5	4.5	520	3	AAG36717	Aag36717 Arabidops
33	164.5	4.5	571	3	AAG36716	Aag36716 Arabidops
34	164.5	4.5	629	3	AAG36715	Aag36715 Arabidops
35	164.5	4.5	2703	4	ABB63299	Abb63299 Drosophil
36	164.5	4.5	2703	5	ABG70019	Abg70019 Larval vi
37	163.5	4.5	522	3	AAG13630	Aag13630 Arabidops
38	163.5	4.5	539	3	AAG13629	Aag13629 Arabidops
39	163.5	4.5	540	3	AAG13628	Aag13628 Arabidops
40	163.5	4.5	821	5	AAM78983	Aam78983 Human pro
41	163.5	4.5	2237	5	ABG70004	Abg70004 Larval vi
42	163.5	4.5	2309	4	ABB66232	Abb66232 Drosophil
43	163.5	4.5	2703	4	ABB60074	Abb60074 Drosophil
44	163	4.4	688	4	ABB67404	Abb67404 Drosophil
45	163	4.4	690	4	ABB66052	Abb66052 Drosophil
46	163	4.4	1128	4	ABB58911	Abb58911 Drosophil
47	163	4.4	3080	4	ABB64877	Abb64877 Drosophil
48	162.5	4.4	2017	6	ABU54584	Abu54584 Human NOV
49	162	4.4	749	4	ABB62915	Abb62915 Drosophil
50	162	4.4	816	2	AAR71111	Aar71111 Spinocere
51	162	4.4	816	2	AAY13494	Aay13494 Human SCA
52	162	4.4	816	6	ABR43117	Abr43117 Human ata
53	161.5	4.4	2062	7	ADC31187	Adc31187 Human nov
54	161	4.4	971	2	AAW48896	Aaw48896 Candida a
55	161	4.4	1322	4	ABB60275	Abb60275 Drosophil
56	161	4.4	1537	4	ABB66989	Abb66989 Drosophil
57	160.5	4.4	1441	4	AAE12570	Aae12570 Drosophil
58	160.5	4.4	1441	5	AAE25569	Aae25569 Human ste
59	160.5	4.4	1441	6	AAE36003	Aae36003 Human SRC
60	159	4.4	993	4	ABB64403	Abb64403 Drosophil
61	159	4.3	3201	4	ABB62899	Abb62899 Drosophil
62	158.5	4.3	459	5	ABP73940	Abp73940 Candida a
63	158.5	4.3	926	4	ABB65135	Abb65135 Drosophil
64	158.5	4.3	1061	2	AAW26370	Aaw26370 Human ste
65	158	4.3	1494	5	AAU78460	Aau78460 Mouse bet
66	158	4.3	3572	5	ABG95659	Abg95659 Human nuc
67	157.5	4.3	296	7	ADD45442	Add45442 Rat Prote
68	157.5	4.3	296	7	ADE57105	Ade57105 Rat Prote
69	157.5	4.3	1566	6	ABR59716	Abr59716 Human RER
70	157	4.3	1424	4	AAM39253	Aam39253 Human pol
71	157	4.3	1586	6	ABU21467	Abu21467 Protein e
72	156.5	4.3	594	4	ABB61362	Abb61362 Drosophil
73	156.5	4.3	700	4	ABB61671	Abb61671 Drosophil
74	156.5	4.3	3238	4	ABB71715	Abb71715 Drosophil
75	156	4.3	531	3	AAB56866	Aab56866 Human pro
76	156	4.3	1703	4	ABB66223	Abb66223 Drosophil
77	155.5	4.2	1367	7	ADB70311	Adb70311 C. neofor
78	155	4.2	392	4	ABG10550	Abg10550 Novel hum
79	155	4.2	404	2	AAR90675	Aar90675 Human chr
80	155	4.2	647	4	AAB67093	Aab67093 Drosophil
81	155	4.2	672	4	ABB59473	Abb59473 Drosophil
82	155	4.2	1153	6	AAG79910	Aag79910 MAML2. 4/
83	155	4.2	1153	7	ADC59310	Adc59310 Human pol
84	155	4.2	1445	4	ABB64619	Abb64619 Drosophil
85	154.5	4.2	1427	4	ABB60419	Abb60419 Drosophil
86	154.5	4.2	1447	5	AAW81029	Aaw81029 Murine pC
87	154.5	4.2	1902	6	ABJ38695	Abj38695 Human nuc
88	154	4.2	530	3	AAG40002	Aag40002 Arabidops
89	154	4.2	581	3	AAG40001	Aag40001 Arabidops
90	154	4.2	639	3	AAG40000	Aag40000 Arabidops
91	153.5	4.2	502	3	AAG24542	Aag24542 Arabidops
92	153.5	4.2	1381	5	ABG93234	Abg93234 Candida a
93	153.5	4.2	1381	5	ABP73947	Abp73947 Candida a
94	153	4.2	274	7	ADD47240	Add47240 Rat Prote
95	153	4.2	274	7	ADD48703	Add48703 Rat Prote
96	153	4.2	1464	4	AAM41039	Aam41039 Human pol
97	152.5	4.1	1101	6	ABR47437	Abr47437 Breast Ca
98	152	4.1	349	4	ABB60321	Abb60321 Drosophil

```
  99  152  4.1   724   4  ABB5778   Abb65778 Drosophil
 100  152  4.1  1012   2  AAY17406  Aay17406 Human atr

                    ALIGNMENTS
```

Db 121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
Qy 181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEBIDLEKSASSHDRVNDYNASLRQESRNRL 240
Db 181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEBIDLEKSASSHDRVNDYNASLRQESRNRL 240
Qy 241 YRDGGKTRLKDTDNGAESHlATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
Db 241 YRDGGKTRLKDTDNGAESHlATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
Qy 301 DVSDDSMVDSISSIDVSPDDVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
Db 301 DVSDDSMVDSISSIDVSPDDVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
Qy 361 IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420
Db 361 IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420
Qy 421 SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480
Db 421 SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480
Qy 481 GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPGNGYFPPYGMMPTIMNPY 540
Db 481 GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPGNGYFPPYGMMPTIMNPY 540
Qy 541 CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600
Db 541 CSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600
Qy 601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTTVTQTTRDGGGVTR 660
Db 601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTTVTQTTRDGGGVTR 660
Qy 661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
Db 661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695

RESULT 1
AAY45097
ID   AAY45097 standard; protein; 695 AA.
XX
AC   AAY45097;
XX
DT   05-JUN-2000   (first entry)
XX
DE   Arabidopsis thaliana early-flowering protein, ELF3.
XX
KW   ELF3; early-flowering; circadian clock; photoperiodism; elf3 mutant;
KW   photoperiod-insensitive; hypocotyl elongation; transgenic plant; corn;
KW   rice; barley; cotton.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200009658-A2.
XX
PD   24-FEB-2000.
XX
PF   17-AUG-1999;   99WO-US018747.
XX
PR   17-AUG-1998;   98US-0096802P.
XX
PA   (UTOR-) UNIV OREGON.
XX
PI   Wagner R, Hicks KA, Spence MZ, Albertson TM, Foss H, Prigge M;
XX
DR   WPI; 2000-224320/19.
DR   N-PSDB; AAZ51052, AAZ51053.
XX
PT   Novel ELF3 gene encoding plant proteins involved in photoperiodism and
PT   circadian rhythms used to alter photoperiodic and/or circadian clock-
PT   based gene expression of plants.
XX
PS   Claim 1a; Page 34-36; 45pp; English.
XX
CC   The present amino acid sequence is the Arabidopsis ELF3 (early-flowering)
CC   protein, that regulates the circadian clock functions and photoperiodism
CC   in plants. The elf3 mutants are photoperiod-insensitive and have long-
CC   hypocotyl, since they are defective in blue-light dependent inhibition of
CC   hypocotyl elongation. The ELF3 protein can complement an elf3 mutant. The
CC   ELF3 sequences are used to alter photoperiodic and circadian clock- based
CC   gene expression in transgenic plants. Such transgenic plants include
CC   Arabidopsis, pepper, tomato, broccoli, wheat, cauliflower, rice, cabbage,
CC   canola, bean, soybean, corn, barley, citrus, cotton, cassava and walnut
CC   having altered levels of ELF3 protein. ELF3 regulatory sequence can be
CC   linked to a foreign gene and used to obtain a circadian clock-based gene
CC   expression
XX
SQ   Sequence 695 AA;

Query Match   100.0%; Score 3669; DB 3; Length 695;
Best Local Similarity 100.0%; Pred. No. 3.3e-289;
Matches 695; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
Db   1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
Qy   61 NTSTIVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
Db   61 NTSTIVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
Qy 121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180

RESULT 2
AAO15964
ID   AAO15964 standard; protein; 695 AA.
XX
AC   AAO15964;
XX
DT   07-FEB-2003   (first entry)
XX
DE   Arabidopsis thaliana nematode feeding site-specific protein #2.
XX
KW   Nematode feeding site-specific; cyst nematode inducible transcription;
KW   regulatory DNA sequence; root knot inducible transcription;
KW   nematode-resistant plant.
XX
OS   Arabidopsis thaliana.
XX
PN   US6448471-B1.
XX
PD   10-SEP-2002.
XX
PF   21-JAN-1999;   99US-00234827.
XX
PR   22-JAN-1998;   98US-0072142P.
XX
PA   (PUZI/) PUZIO P S.
PA   (GRUN/) GRUNDLER F M W.
XX
PI   Puzio PS, Grundler FMW;
XX
DR   WPI; 2003-038247/03.
DR   N-PSDB; AAL50923.
XX
PT   New isolated regulatory DNA sequences obtained from Arabidopsis thaliana,
PT   useful for promoting root knot and cyst nematode-inducible transcription

```
of an associated DNA sequence.
PT
XX
PS   Disclosure; Col 39-42; 29pp; English.
XX
CC   The invention comprises an Arabidopsis thaliana nematode feeding site-
CC   specific regulatory DNA sequence that promotes root knot and cyst
CC   nematode inducible transcription. The Arabidopsis thaliana DNA sequence
CC   of the invention is useful in promoting root knot and cyst nematode
CC   inducible transcription of an associated DNA sequence when re-introduced
CC   into a plant. The Arabidopsis thaliana DNA sequence is useful in making
CC   plants that are resistant, or at least less susceptible to plant
CC   parasitic nematodes. The present amino acid sequence represents an
CC   Arabidopsis thaliana nematode feeding site-specific protein
XX
SQ   Sequence 695 AA;

Query Match             90.7%;  Score 3327;  DB 6;  Length 695;
Best Local Similarity   91.2%;  Pred. No. 2.1e-261;
Matches   649;  Conservative   4;  Mismatches   25;  Indels   34;  Gaps    3;

Qy    1  MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    1  MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTRNSRSN  60

Qy   61  NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120

Qy  121  REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180

Qy  181  ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240
         |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  181  ATCSKPEVRDQVKANRRSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240

Qy  241  YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300

Qy  301  DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  301  DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360

Qy  361  IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPLPHVVVKQRGDSEKTDQHKMES 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPLPHVVVKQRGDSEKTDQHKMES 420

Qy  421  SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCPPQPPPSGNHQ----------- 469
         |||||||||||||||||                                  :
Db  421  SAENVVGRLSNQGHHQQHPTTCLF------------------QQTTHRLHRLQMDIAFLLSL 463

Qy  470  --QWLIPVMSPSEGLI----YKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPG 523
            :   :                 ||||||||||||||||||||||||||||||||||||
Db  464  LLQEIISNGDPCNVSLGRTDIQPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPG 523

Qy  524  NGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQ 583
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  524  NGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQ 583

Qy  584  QQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPPAAVDEDSNINNAPEQTMTTTTTT 643
         |||||||||||||||||||||||||||||||||||  ||||||||||||||||||||||
Db  584  QQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFGPFAAVDEDSNINNAPEQTMTTTTTT 643

Qy  644  TRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  644  TRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695

RESULT 3
AAO15963
ID   AAO15963 standard; protein; 695 AA.
XX
```

```
AC   AAO15963;
XX
DT   07-FEB-2003  (first entry)
XX
DE   Arabidopsis thaliana nematode feeding site-specific protein #1.
XX
KW   Nematode feeding site-specific; cyst nematode inducible transcription;
KW   regulatory DNA sequence; root knot inducible transcription;
KW   nematode-resistant plant.
XX
OS   Arabidopsis thaliana.
XX
PN   US6448471-B1.
XX
PD   10-SEP-2002.
XX
PF   21-JAN-1999;  99US-00234827.
XX
PR   22-JAN-1998;  98US-0072142P.
XX
PA   (PUZI/) PUZIO P S.
PA   (GRUN/) GRUNDLER F M W.
XX
PI   Puzio PS, Grundler FMW;
XX
DR   WPI; 2003-038247/03.
DR   N-PSDB; AAL50922.
XX
PT   New isolated regulatory DNA sequences obtained from Arabidopsis thaliana,
PT   useful for promoting root knot and cyst nematode-inducible transcription
PT   of an associated DNA sequence.
XX
PS   Disclosure; Col 31-34; 29pp; English.
XX
CC   The invention comprises an Arabidopsis thaliana nematode feeding site-
CC   specific regulatory DNA sequence that promotes root knot and cyst
CC   nematode inducible transcription. The Arabidopsis thaliana DNA sequence
CC   of the invention is useful in promoting root knot and cyst nematode
CC   inducible transcription of an associated DNA sequence when re-introduced
CC   into a plant. The Arabidopsis thaliana DNA sequence is useful in making
CC   plants that are resistant, or at least less susceptible to plant
CC   parasitic nematodes. The present amino acid sequence represents an
CC   Arabidopsis thaliana nematode feeding site-specific protein
XX
SQ   Sequence 695 AA;

Query Match             90.5%;  Score 3321;  DB 6;  Length 695;
Best Local Similarity   91.0%;  Pred. No. 6.4e-261;
Matches   648;  Conservative   4;  Mismatches   26;  Indels   34;  Gaps    3;

Qy    1  MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db    1  MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTRNSRSN  60

Qy   61  NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120

Qy  121  REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180

Qy  181  ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240
         |||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db  181  ATCSKPEVRDQVKANRRSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240

Qy  241  YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300

Qy  301  DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
         |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||
Db  301  DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIAKQQRVFAVQLFELHRLIKVQKL 360
```

```
CC   that is differentially expressed in neuronal tissue of a first animal
CC   subjected to pain, a method for identifying a compound which regulates
CC   the expression of a polynucleotide sequence which is differentially
CC   expressed in an animal subjected to pain, a method for identifying a
CC   compound that regulates the activity of one or more of the
CC   polynucleotides, a method for producing a pharmaceutical composition, a
CC   method for identifying a compound or small molecule that regulates the
CC   activity in an animal of one or more of the polypeptides given in the
CC   specification, a method for identifying a compound useful in treating
CC   pain and a pharmaceutical composition comprising the one or more
CC   polypeptides or their antibodies. The polynucleotide or the compound that
CC   modulates its activity is useful for preparing a medicament for treating
CC   pain (e.g. spinal segmental nerve injury (Chung), chronic constriction
CC   injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene
CC   therapy). The sequence presented is a human protein (shown in Table 2 of
CC   the specification) which is differentially expressed during pain. Note:
CC   The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic form directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 971 AA;

Query Match             5.4%;  Score 197;  DB 7;  Pred. No. 1e-06;
Best Local Similarity   19.4%;
Matches  142;  Conservative   94;  Mismatches  245;  Indels  250;  Gaps   33;

Qy      77  VERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRR  136
                                                         ::    :::    ::  :
Db      93  LSRDSSQEYTDSTG------------IDLHEFLVNTLKNNPRDRMMLLKLEQE--ILDFIGNNE  142

Qy     137  SQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEV---------NQTGSKQNVCLATCSKPEV  188
               :          : ::   :   :::  : :          ::  :   :    : ::
Db     143  S--------------PRKKFPPM--TSYHRMLLHRVAAYFGLDHNVDQSGKSVIVNKTSNTRI  189

Qy     189  RDQV---------KANARSGGFVISLDVSVTEEIDLE----------KSASSHDRVNDYNAS  231
               :          :    :            :::           :::  :  :
Db     190  PDQKFNEHIKDDKGEDFQKRYILKRDNSSFDKDDNQVRIRLKDDRRSKSIEEREEEY----  246

Qy     232  LRQESRNRLYRDGG---------KIRLKDTD------------NGAESHLATENHSQEG  269
               :    ::  : ::          ::  :                ::    :::
Db     247  --QRARDRIFSQDSLCSQENYIIDKRLQDEDASSTQQRRQIFRVNKDASGRSTNSHQ---  301

Qy     270  HGSPEDIDNDRRYSKSRRACASLQQINEEASDDVSDDSM--------VDSISSIDV-SPD  319
                     :   :  :          :          :         :: ::: :
Db     302  ---SSTENELKYSEPRPWSSTD------------SDSSLRNLKPAVTKASSFSGISVLTRG  347

Qy     320  DVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLDEISFLGKVS  379
                 :  :: ::                                        :     :
Db     348  DSSG---------------SSKSIG-------------RLSK----TGQPFINPDGSPVVVNPP  379

Qy     380  AKSYPVKKLLPSEFLVKPLP----------------HVVVKQRGDSEKTDQH  416
                 :       ::   : :: ::: :          ::  :  :
Db     380  MTQQPVRSQVEGP--PQPPLPAPPQQPAANHIFSQQDNLGSQFSHMSLARQPSADGSDPH  437

Qy     417  KMESSAENVVGRLSNQGHHQQSNYMPPANNPPASPAPNGYCFPPQPP----------  464
                :              ::                        :
Db     438  AAMFQSTVVL-------QSPQQSGYIMTAAPPPHPPPP----PPPPPPPPLPPGQPVPT  486

Qy     465  ----SGNHQQWLIPVMSP----SEGLIYKPHPGM------AHTGHYGGYYGHYMPTPMVMPQ  512
                :                  :    :  : :              ::
Db     487  AGYPASGH---------PVSQPVLQQQGYIQQPSPOMPACYCAPGHYHSSQPQYRPVPSV--H  538

Qy     513  YHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ----QQPNEQMNOFGHPGNLQN-  566
               :     :      :                         ::
Db     539  YNSHLNQPLPQPAQQTGYQVIP-------NQQQNYQGIVGVGVQQPQSQLVSGQPNSIGNQ  591

Qy     567  -------TQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGS  614
                      :  ::                        :  :
Db     592  IQGVVIPYTSVPIYQVSLPQGSQGIPHQTYQQPVMFP----NQSNQGSMPT---TGM  641

Qy     615  KSFRPFAAVDEDSNINNA-----------PEQTMTTTTTT------------RTTVTQTTRDGGG  657
              :::::                       :::::::::
Db     642  PVYYSVIPPGQQNNLSSSVGYLQHPGSEQVQPFRTTSPCSSQQLQHQCTAGPPPPPGGG  701
```

```
Qy     361  IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES  420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     361  IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES  420

Qy     421  SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPSGNHQ----------  469
              |||||||||||||||||||                              ||||||
Db     421  SAENVVGRLSNQGHHQHPTCLF----------QQTTHRLHRLQMDIAFLLSL          463

Qy     470  --QWLIPVMSPSEGLI---YKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPG  523
                        :  ||;:|   ||||||||||||||||||||||||||||||||||||||
Db     464  LLQEIISNGDPCNVSLGRTDIQPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPG  523

Qy     524  NGYFPPYGNMPTIMNPYCSSQQQQQQPNEQMNOFGHPGNLQNTQQQQRSDNEPAPQQQ  583
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     524  NGYFPPYGNMPTIMNPYCSSQQQQQQPNEQMNOFGHPGNLQNTQQQQRSDNEPAPQQQ  583

Qy     584  QQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTT  643
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     584  QQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFGPFAAVDEDSNINNAPEQTMTTTTTT  643

Qy     644  TRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP  695
              |||||||||||||||||||||||||||||||||||||||||||||||||||
Db     644  TRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP  695
```

```
RESULT 4
ADE56304
ID   ADE56304 standard; protein; 971 AA.
XX
AC   ADE56304;
XX
DT   29-JAN-2004  (first entry)
XX
DE   Human Protein Q15032, SEQ ID NO 2156.
XX
KW   Human; pain; neuronal tissue; gene therapy;
KW   spinal segmental nerve injury; chronic constriction injury; CCI;
KW   spared nerve injury; SNI; Chung.
XX
OS   Homo sapiens.
XX
PN   WO2003016475-A2.
XX
PD   27-FEB-2003.
XX
PF   14-AUG-2002; 2002WO-US025765.
XX
PR   14-AUG-2001; 2001US-0312147P.
PR   01-NOV-2001; 2001US-0346382P.
PR   26-NOV-2001; 2001US-0333347P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
PA   (FARB ) BAYER AG.
XX
PI   Woolf C, D'urso D, Befort K, Costigan M;
XX
DR   WPI; 2003-268312/26.
DR   GENBANK; Q15032.
XX
PT   New composition comprising two or more isolated polypeptides, useful for
PT   preparing a medicament for treating pain in an animal.
XX
PS   Claim 1; Page; 1017pp; English.
XX
CC   The invention discloses a composition comprising two or more isolated rat
CC   or human polynucleotides or a polynucleotide which represents a fragment,
CC   derivative or allelic variation of the nucleic acid sequence. Also
CC   claimed are a vector comprising the novel polynucleotide, a host cell
CC   comprising the vector, a method for identifying a nucleotide sequence
CC   which is differentially regulated in an animal subjected to pain and a
CC   kit to perform the method, an array, a method for identifying an agent
CC   that increases or decreases the expression of the polynucleotide sequence
```

```
Qy      81  LSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKM--------VREEEDFAVPV  130
            ||||  :: ::|||.|
Db     253  LSVPHDSTISTTASASATSGARSNDQRKPPLSDAQRRMRFPRADLSKPITEEEHDRYAA  312

Qy     131  YINSRRSQSHGRTK--------------------SGIEKEKHTPMVAPSSHHSIRFQEVN  170
            |:                              : ||:
Db     313  YL--------HGENKITEMHNIPPKSRLFIGNLPLKNVSKEDLFRIFSPYGH------IM  358

Qy     171  QTGSKQNVCLATCSKPE-VRDQVKANARSGFVISLDVSVTEEIDEKSASS-------H  222
            ||: : ::|||.|:
Db     359  QINIKNAFGFIQFDNPQSVRDAIECE-------SQEMNFGKKLILEVSSSNARPQFDHG  410

Qy     223  DRVNDYNASLRQESRNRLYRDGGKTRLKDTDNG---------------AESHLAT  262
            :||:
Db     411  DHGTNSSSTFISSAKRPFQTESGD-MYNDDNGAGYKKSRRHTVSCNIFVKRTADRTYAI  468

Qy     263  B--NHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDD  320
            :  :.|
Db     469  EVFNRFRDGTG---LETDMIFLKPRMELG-KLINDAAYNGVWGVVLVNKTHNVDV-----  519

Qy     321  VVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSA  380
                                          :|||:
Db     520  ---------------QTFYKGSQGET-----------KFDEYISISAD--DAVAIFNNIK-  551

Qy     381  KSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNY  440
                                                                :|||
Db     552  -----------------------------------NNRNNSRPTDYRAMS--------HQQNIY  572

Qy     441  ---MPFANNPPASPAP------NGYCFPP----QPPPSGNHQQWLIPVMSPSEGLIYKPH  487
               ||||
Db     573  GAPPLPVPNGPAVGPPPQTNYYQGYSMPPPQQQQQQPYGNYG--MPPPSHDQG--YGSQ  627

Qy     488  PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGF------PPPGNGYFPP----  529
            |||
Db     628  PPIPMNQSYGRYQTSIPPPP--PQQQIPQGYGRYQAGPPPQ---PPSQTPMDQQQLLS  680

Qy     530  --YGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPT  587
                                                   :|:         ||||||||
Db     681  AIQNLPPNVVSNLLSMAQQQQQQPHAQQQLV------GLIQSMQGQ------APQQQQQQL  729

Qy     588  KSYPPRARKSRQGSTGSSPSGPQGISGSKSFRPFA  621
                    :: :|:|.|.| :|
Db     730  GGYSSMNSSSPPPMSTNYNG-QNISAKPSAPPMS  762
```

RESULT 5
ABG93215
ID   ABG93215 standard; protein; 802 AA.
XX
AC   ABG93215;
XX
DT   21-NOV-2002 (first entry)
XX
DE   S. cerevisiae BAX-associated protein fragment SEQ ID 388.
XX
KW   Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
KW   vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
KW   apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW   neurodegeneration; cell death.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO200264766-A2.
XX
PD   22-AUG-2002.
XX
PF   21-DEC-2001; 2001WO-EP015398.
XX
PR   22-DEC-2000; 2000EP-00870318.
PR   04-JAN-2001; 2001EP-00870002.
PR   09-JAN-2001; 2001EP-00870003.
XX
PA   (JANC) JANSSEN PHARM NV.
XX
PI   Contreras RH, Eberhardt I, Luyten WHML, Reekmans RJ;
XX
DR   WPI; 2002-667002/71.
DR   N-PSDB; ABQ76481.
XX
PT   New isolated nucleic acid representing a synthetic BAX-gene, useful as
PT   medicament for treating, preventing and/or alleviating yeast or fungal
PT   infections or proliferative disorders, or for preventing apoptosis in
PT   certain diseases.
XX
PS   Claim 36; Fig 1; 344pp; English.
XX
CC   This invention describes a novel nucleic acid representing a synthetic
CC   Bax gene. The Bax gene of the invention is useful for identifying Bax-
CC   resistant yeast or fungi, identifying, or obtaining and identifying
CC   Candida spp. sequences that are differentially expressed in a pathway
CC   eventually leading to programmed cell death or identifying inhibitors or
CC   inhibitor sequences of Bax-induced cell death. The products of the
CC   invention have cytostatic, fungicide, immunosuppressive, virucide and
CC   vasotropic activity and can be used in vaccines or for gene therapy. The
CC   isolated nucleic acids, polypeptides, pharmaceutical compositions,
CC   antisense molecules and antibodies are useful as medicaments or in
CC   preparing a medicament for treating, preventing and/or alleviating
CC   diseases associated with yeast or fungi or proliferative disorders, such
CC   as cancer, or for preventing apoptosis in certain diseases. The compounds
CC   or polypeptides, or the genetically modified organism are useful for
CC   preparing a medicament for modifying the endogenic flora of humans and
CC   other mammals. The vaccine is useful for immunising against yeast or
CC   fungal infections. Apoptosis-related diseases include autoimmune disease,
CC   ischaemia, diseases related with viral infections or neurodegenerations.
CC   This sequence represents a polypeptide associated with the Bax gene
CC   described in the disclosure of the invention
XX
SQ   Sequence 802 AA;

Query Match   5.3%; Score 193.5; DB 5; Length 802;
Best Local Similarity 20.2%; Pred. No. 1.5e-06;
Matches 128; Conservative 73; Mismatches 216; Indels 217; Gaps 29;

```
Qy     658  VTRVIKVVPHN  668
            :: .: ||:|
Db     702  MVMMQLSVPNN  712
```

RESULT 6
ABR53105
ID   ABR53105 standard; protein; 802 AA.
XX
AC   ABR53105;
XX
DT   20-JUN-2003 (first entry)
XX
DE   Protein sequence #SEQ ID 1075.
XX
KW   Multiprotein complex; eukaryote; drug target; diagnosis.
XX
OS   Saccharomyces cerevisiae.
XX
PN   EP1258494-A1.
XX
PD   20-NOV-2002.
XX
PF   20-DEC-2001; 2001EP-00130253.
XX
PR   15-MAY-2001; 2001EP-00111774.
XX
PA   (CELL-) CELLZOME AG.
XX
PI   Bauer A, Gavin A, Grandi P, Krause R, Kruse UD, Kuester BD;
PI   Marzioch M, Schultz JD, Superti-Furga GD;
XX
```

```
DR   WPI; 2003-250078/25.
DR   N-PSDB; ACC61147.
XX
PT   New isolated protein complexes useful for diagnosing a disease or
PT   disorder, or as a target for an active agent of a pharmaceutical,
PT   preferably a drug target in the treatment or prevention of disease or
PT   disorder.
XX
PS   Disclosure; SEQ ID NO 1075; 17pp + Sequence Listing; English.
XX
CC   The invention relates to multiprotein complexes from eukaryotes. Proteins
CC   of the invention and DNA sequences encoding them are given in records
CC   ABR52568-ABR53903 and ACC60610-ACC61944 respectively. The complexes are
CC   obtainable by using a protein as a bait and isolating the set of proteins
CC   which is attached thereto from cells. Such protein complexes may comprise
CC   up to 30 distinct proteins. Protein complexes of the invention are useful
CC   for diagnosing a disease or disorder, or as a target for an active agent
CC   of a pharmaceutical, preferably a drug target in the treatment or
CC   prevention of a disease or disorder. Note: The sequence data for this
CC   patent is not represented in the printed specification, but is based on
CC   sequence information supplied by the European Patent Office. The complete
CC   document is available on CD-ROM
XX
SQ   Sequence 802 AA;

Query Match             5.3%;   Score 193.5;  DB 6;  Length 802;
Best Local Similarity   20.2%;  Pred. No. 1.5e-06;
Matches  128;  Conservative   73;  Mismatches  216;  Indels  217;  Gaps   29;

Qy    81  LSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKM----------VREEEDFAVPV  130
              ||:|||                                         :|| |||
Db   253  LSVPPHDSTISTTASASATSGARSNDQRKPLSDAQRRMFPPRADLSKPITEEEHDRYAA   312

Qy   131  YINSRRSQSHGRTK-------------------SGIEKEKHTPMVAPSSHHSIRFQEVN   170
          |:|                                ||::|||
Db   313  YL------HGENKITEMHNIPPKSRLFIGNLPLKNVSKEDLFRIFSPYGH--------IM  358

Qy   171  QTGSKQNVCLATCSKPE-VRDQVKANARSGGFVISLDVSVTEEIDLEKSASS-------H  222
             ||  ::              :|: |||
Db   359  QINIKNAFGFIQFDNPQSVRDAIECE--------SQEMNFGKKLILEVSSSNARPQFDHG  410

Qy   223  DRVNDYNASLRQESRNRLYRDGGKTRLKDTDNG--------------------ABSHLAT  262
          | :  : |  ||                                        ||  |
Db   411  DHGTNSSSTFISSAKRPFQTESGD--MYNDDNGAGYKKKSRRHTVSCNIFVKRTADRTYAI  468

Qy   263  E--NHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDVSDDSMVDSISSIDVSPDD  320
             :|:                                          : :|:
Db   469  EVFNRFRDGTG----LETDMIFLKPRMELG-KLINDAAYNGWGWVLVNKTHNVDV-----  519

Qy   321  VVGILGQKRFWRARKAIANQQRVFAVQLFELHRLLRKVQKLIAASPDLLLDEISFLGKVSA  380
                                         :| :||                     |:
Db   520  ----------QTFYKGSQET-----------------KPDEYISISAD--DAVAIFNNIK-  551

Qy   381  KSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNY  440
                                                           |||
Db   552  ---------------------------NNRNNSRPTDYRAMS----------HQQNIY   572

Qy   441  ---MPFANNPPASPAP-----NGYCPPP----QPPPSGNHQQWLIPVMSPSEGLIYKPH  487

Db   573  GAPPLPVPNGFAVGPPGPPQTNYYQGYSMPPPQQQQQQPYGNYG---MPPPSHDQG--YGSQ  627

Qy   488  PCMAHTGHYGYGYYGHYMPTPMVMPQYHPGMGF---------PPPGNGYFPP---------  529

Db   628  PPIPMNQSYGRYQTSIPPPP---PQQQIPQGYRYQAGPPPQ---PPSQTPMDQQQLLS  680

Qy   530  --YGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPT  587

Db   681  AIQNLPPNVVSNLLSMAQQQQQQPHAQQQLV---GLIQSMQQO--------APQQQQQQL  729

Qy   588  KSYPRARKSRQGTGSSPSGPQGISGSKSFRPPA  621

Db   730  GGYSSMNSSSPPPMSTNVNG-QNISAKPSAPPMS  762
```

```
RESULT 7
ABB60944
ID   ABB60944 standard; protein; 1503 AA.
XX
AC   ABB60944;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 9624.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL05047.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 9624; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL305111, expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1503 AA;

Query Match             5.2%;   Score 192;  DB 4;  Length 1503;
Best Local Similarity   22.9%;  Pred. No. 4.7e-06;
Matches  151;  Conservative   74;  Mismatches  244;  Indels  190;  Gaps   35;

Qy     4  GKDEEKILEPMFP-----RLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSR   58
          | ::|  |:|                             :                  :
Db   398  GMDGNRRMME--LPQSYDYRVHQYD------RFHPLNQ----RQMS---NVGDY--ER   437

Qy    59  SNNTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRK  118
          ||                     :||
Db   438  SNR------MPLDLDRFKKREPHSESSFLEAPNELSLGQLSVEEPVLKS---EDEN   483

Qy   119  MVREEEDFAVPVVINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV  178
          |                                                 :
Db   484  ISNREMAMKIAV-LDIKDKALHSTGDGESKEKDT----------LEEVKLEGLHEEL   530

Qy   179  CL-ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQES  236
                 : |                              :
Db   531  KLDEESLVKPESKLDSEVSCINGKF------EDRSEMEVENKPLGPVK-REAKAELINQM  583

Qy   237  RNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSK-SRACASLQQIN  295
          :   | :   |  ||            ||
Db
```

```
CC   inhibitor sequences of Bax-induced cell death. The products of the
CC   invention have cytostatic, fungicide; immunosuppressive, virucide and
CC   vasotropic activity and can be used in vaccines or for gene therapy. The
CC   isolated nucleic acids, polypeptides, pharmaceutical compositions,
CC   antisense molecules and antibodies are useful as medicaments or in
CC   preparing a medicament for treating, preventing and/or alleviating
CC   diseases associated with yeast or fungi or proliferative disorders, such
CC   as cancer, or for preventing apoptosis in certain diseases. The compounds
CC   or polypeptides, or the genetically modified organism are useful for
CC   preparing a medicament for modifying the endogenic flora of humans and
CC   other mammals. The vaccine is useful for immunising against yeast or
CC   fungal infections. Apoptosis-related diseases include autoimmune disease,
CC   ischaemia, diseases related with viral infections or neurodegeneration.
CC   This sequence represents a polypeptide associated with the Bax gene
CC   described in the disclosure of the invention
XX
SQ   Sequence 738 AA;

Query Match             5.2%;  Score 189.5; DB 5;  Length 738;
Best Local Similarity  17.5%;  Pred. No. 2.8e-06;
Matches 128;  Conservative 99;  Mismatches 248;  Indels 257;  Gaps    29;

Qy       19 HVN-DADKGGPRAPPRNKMALYEQLSIPSQRFGDHGT-MNSRSNNTSTLVHPGPSSQPCG     76
              ||| ||                            |: :::::::::::
Db      142 HVNLDKGKHVPSKPVSNTTSWAAAVSVDTK----HDVPQDSNDNNNEELEAQGQQAQEKN   197

Qy       77 VERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEEDFAVPVVINSRR   136
                                                                  ::|::   LSSKK  229
Db      198 QEKEQEQQQQEGHNNKEEHKQIEQPS-                                  229

Qy      137 SQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANA   196
                                            :::
Db      230 TTS--RTSASQPKKMSWAAIATPKPKAVKKTE---SPLENVAEL--KKEISDIKK---    277

Qy      197 RSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGA   256
                                        |||:|: ::
Db      278 ------------------------DDQKSEASEEKVNEQETSAQEQEEE-            302

Qy      257 ESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDV   316
                                          :|: |
Db      303 ------TAEPSEENEDRVPEVDGEEVQEEAE--KKEQVKEEQTAEELEQEQDNVA-       350

Qy      317 SPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLG   376
             :|::        |||              VEISAVISEPPE--DQANTVP  381
Db      351 APEEEVTVVEEK-                                                 381

Qy      377 KVSAKS------------------------YPVKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVV   426
               :::
Db      382 QPQQQSQQPQQPQQPQQPQQ-QQQPQQPQQPQQPQQQQQQQPVQAQAQAQE            440

Qy      427 GRLS------------------NQGHHQQSNYM-----PFANNPPASPAPNGY---CFPPQPPPS   465
             :||                                        :|: |||||||
Db      441 EQLSQNYYTQQQQQQYAQQQHQLQQQYLSQQQYAQQQQQHPQPQSQQPQSQPQSQKQ    500

Qy      466 GNH---QQWLI-                           -PVMSPSEGLIY-          484
             :|:                                    VEISAVISEPPE-
Db      501 GNNVAAQQYYMYQNQFPGYSYPGMFDSQGYAYGGQOYQQLAQNNAQTSGNANQYNFQQGYG   560

Qy      485 -KPHPGMAHT-------GHYGG---YYGHY--MPTPMVMPQYHP   515
                                        |||
Db      561 QAGANTAAANLTSAAAAAAASPATAHAQPQQQQPYGGSFMPYYAHFFYQQSFPYGQPYGV   620

Qy      516 GMGFP--PPGNGYFPPYGMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQR   573
                                                     :: ||||||
Db      621 AGQYPYQLPKNNY---------NYYQTQNGQEQQSPNQGVAQHSEDSQKQSQQQQQ     669

Qy      574 ------------SDNEPA-PQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPF   620
Db      670 QPQGQPQPEVQMQNGQPVNPQQQMFQQYYQFQQQQQAAAAAQQGV------PY   722

Qy      621 AAVDEDSNINNA   632
             :|:
Db      723 GYNGYDYNSKNS   734
```

```
Db      584 GNKLKNDGSTLVAK--SSAESPIKEDVPPNLARRSELNVSTGTDLSAVSKKQNSLKRKK   640
Qy      296 EEASDDVSD--DSMVD--------SISSIDVSPDDVVGILGQKRFWRARKAIANQQRV   343
Db      641 SSATSDDSDKHDSNPDCDTECKSEDAEKSETDAEKDD-----------           677
Qy      344 FAVQLFELHRLIKVQKLIAASPDLLLDEIS-FLGKVSAKSYP------VKKLLPSEFLVK   396
Db      678 ---SIVKVNIKLKLQEPVEEKSE--CEEISKVVGKGAAASAPKVNPDAVVKLLANELSMD   732
Qy      397 PPLPHVVVKQRG-----------------DSEKTDQHKMESSAENVVGRLS--NQGHH   435
Db      733 ----------KDRGKRQARHRYRQNRSKRSLEDNDQHESDDKQHAVSKRLAPLLDDFED   784
Qy      436 QQSNYMPFANNPPASPAPNGYCFPQPPPSGNHQQWLIPVMSPSEGLIYKPHPGMAH---   492
Db      785 HNGNRVGFGN------LGNG-MLTPDAEYSGERS-------------YPSHEMLAHFDE   823
Qy      493 --TGHYG-GYYGHYMPTPMVMPQVH--PGMGFPPGNGYFPPYGMPTIMNPYCSSQQQQ   547
Db      824 LETAHDGDSMVPIYDPTKVVVPMVNQRPNGF--------QGQQPVQMQ--FSQQQP   870
Qy      548 QQQPNEQMNQFGH----PGNLQNTQQQQQRSDNEPAPQQQQ-QPTKSYPRARKSRQGS   600
Db      871 QQQMQQQMQQ-QHMQQQPPQQQQPQQQQQQQQQQQQQQQVQQPQQHDLSRQRHQSS     928
```

```
RESULT 8
ABG93140
ID   ABG93140 standard; protein; 738 AA.
XX
AC   ABG93140;
XX
DT   21-NOV-2002 (first entry)
XX
DE   S. cerevisiae BAX-associated protein fragment SEQ ID 238.
XX
KW   Bax; Bax-resistance; cytostatic; fungicide; immunosuppressive; virucide;
KW   vasotropic; vaccine; gene therapy; proliferative disorder; cancer;
KW   apoptosis; fungal; yeast; infection; autoimmune disease; ischaemia;
KW   neurodegeneration; cell death.
XX
OS   Saccharomyces cerevisiae.
XX
PN   WO20264766-A2.
XX
PD   22-AUG-2002.
XX
PF   21-DEC-2001; 2001WO-EP015398.
XX
PR   22-DEC-2000; 2000EP-00870318.
PR   04-JAN-2001; 2001EP-00870002.
PR   09-JAN-2001; 2001EP-00870003.
XX
PA   (JANC ) JANSSEN PHARM NV.
XX
PI   Contreras RH, Eberhardt I, Luyten WHML, ReeKmans RJ;
XX
DR   WPI; 2002-667002/71.
DR   N-PSDB; ABQ76406.
XX
PT   New isolated nucleic acid representing a synthetic BAX-gene, useful as
PT   medicament for treating, preventing and/or alleviating yeast or fungal
PT   infections or proliferative disorders, or for preventing apoptosis in
PT   certain diseases.
XX
PS   Claim 36; Fig 1; 344pp; English.
XX
CC   This invention describes a novel nucleic acid representing a synthetic
CC   Bax gene. The Bax gene of the invention is useful for identifying Bax-
CC   resistant yeast or fungi, identifying, obtaining and identifying
CC   Candida spp. sequences that are differentially expressed in a pathway
CC   eventually leading to programmed cell death or identifying inhibitors or
```

```
Qy   155  MVAPSHHSIRFQEVNQTGSK---QNVCLATCSKPEVRDQVKANARSGGFVISLLDVSVTE  211
               :|||  :| ::   :  :  : :  ::| :|  : ||
Db     6  MSAPPSIPITTPGTSSNTRRVARRACLSCREK----KIKCNGEPMTTITAAD--GTN      58

Qy   212  EIDLEKSAS-SHDRVNDYNASLRQESRNRLYRDGGKTRLKDTD---NGAESHLATENHSQ  267
              |||:|::   | :|    |:||                    ||| |:
Db    59  RIIPEKTRTCSNCRFGIGIPCVFVQSNR------GGKRKKRSTEHQLSDSEATTAPATH-- 110

Qy   268  EGHGSPEDIDNDRE--YSKSRACALSQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI   324
          ||| :     :  |||   |:        :     ||      :
Db   111  -----PNQLGQHQQLPPIKTKSTPSIHNL-------DSGVV-IPSI-------         144

Qy   325  LGQKRFWRARKAIANQQRVFAVQLFELHRLLKVQKLIAASPDLLLDEISFLGKVSAKSYP  384
          ||:   :: :   :|||      : :          || ::    : :
Db   145  -----NQQQERHQDGYRNEELASTERV---SSYILD-----GNNSTK----         178

Qy   385  VKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSA-------ENVVGRLS---     430
            |    :      ::  :   : :|: :::: :: ::|||      :||
Db   179  ---LPS-----PVLTHTTFDSRRDEGQRFNNQLRRSSTTTAGFSLPPPQELLGRRSVSY  229

Qy   431  NQGHHQQSNYMPF-------ANNPPASPAPNGYCFPPQPP------PSGNHQQWLI      473
          : |:|||:: |        ||     :        |:||||      | :|||||:|
Db   230  SPGPSSIKSQLPHLTSSSTTTSSVQSPPPPPP------PPQPPRGMGIQFHEASHQPGIF  283

Qy   474  PVMSPSEGLIYKPHPGMAHTCH-YGGYYGHYMPTP---------MVMPQYHPGMGFPP   521
             ||    |||| : |:      |:| :: | |          :|  ||| |: ||||
Db   284  PPPPP------PPPPPPPHPDHLHHQYYGYPPPPPPPPPPLHHHHHYPPSHHPGFGFPP   337

Qy   522  PGNGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPG--NLQNTQQQQQR---     573
          | || |||||  |  ||  |:|  |: ||::  :|| || ||     |::::|:|
Db   338  PPPGPPGPPGPPPVPPPHYHQSAPPESPPKEFKHR--HPKYFHKQNNKGHRHRHHPHLH  395

Qy   574  SDNEPAPQQ--------QQQPTKSYPRARKSRQGSTGSSPSGPQ----GISGSKSFR--  618
            ::| |||        |||  ||:   |:| ||: |:|| ||       |:|||||:
Db   396  HHHPPPPHHHHHPRDFQQHNGQDDIKSKHTDENSQNSKPSSPRWLDNMSVSSSDSRSV   455

Qy   619  --PFRAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENA 676
            |: || :: :|: ::::  :  ||||||| : :|||     || |:|
Db   456  SSPQSQHNDKSQYSNHKSSTQMSSCT--NSVTQT--LDRLPKIVSTQQNNLIPTS      506

Qy   677  ARIFQS--IQERKRYDSSK        694
          || |:|||  |:|||| |:|
Db   507  KILSNSPFTESELAKYDLPK       526
```

RESULT 9
ABP73921
ID   ABP73921 standard; protein; 1130 AA.
XX
AC   ABP73921;
XX
DT   30-JAN-2003   (first entry)
XX
DE   Candida albicans essential protein SEQ ID NO 7758.
XX
KW   Fungus; yeast; tetracyclin; promoter; GRACE strain; biosynthesis;
KW   signal transduction; DNA replication; cell division; growth;
KW   proliferation; Candida albicans; fungicide; antifungal.
XX
OS   Candida albicans.
XX
PN   WO200253728-A2.
XX
PD   11-JUL-2002.
XX
PF   26-DEC-2001; 2001WO-US049486.
XX
PR   29-DEC-2000; 2000US-0259128P.
PR   20-FEB-2001; 2001US-00792024.
PR   22-AUG-2001; 2001US-0314050P.
XX
PA   (ELIT-) ELITRA PHARM INC.
XX
PI   Roemer T, Jiang B, Boone C, Bussey H, Ohlsen KL;
XX
DR   WPI; 2002-566694/60.
DR   N-PSDB; ABZ32471.
XX
PT   Constructing strains for identifying gene products as effective targets
PT   for therapeutic intervention, by inactivating in the strain one allele of
PT   a gene and placing other allele of the gene under conditional expression.
XX
PS   Claim 44; SEQ ID NO 7758; 167pp + Sequence Listing; English.
XX
CC   The invention relates to constructing (M1) a strain of diploid fungal
CC   cells in which both alleles of a gene are modified, comprising modifying
CC   one allele by insertion or replacement by a cassette having an
CC   expressible selectable marker and modifying other allele by
CC   recombination, of a promoter replacement fragment with a heterologous
CC   promoter, so that expression of the second allele is regulated by the
CC   promoter. (M1) is useful for constructing a strain of diploid fungal
CC   cells in which both alleles of a gene are modified. The diploid fungal
CC   cells having both alleles modified are useful for identifying a gene that
CC   is essential to the survival or growth of a fungus, a gene that
CC   contributes to the virulence and/or pathogenicity of a fungus, a gene
CC   that contributes to the resistance of a diploid fungus to an antifungal
CC   agent, an antifungal agent that inhibits the growth of a diploid fungus
CC   and for identifying a therapeutic agent for treatment of a mammalian
CC   disease. (M1) is useful for identifying a compound which modulates the
CC   activity of a gene product, preferably enzymatic activity, carbon
CC   compound catabolism, biosynthetic, transporter, transcriptional,
CC   translational, signal transduction, DNA replication and cell division
CC   activity. The method is useful for identifying a compound having the
CC   ability to inhibit growth or proliferation of a fungus, a compound having
CC   treating infection by C. albicans. The present sequence is that of an
CC   essential Candida albicans protein used in the method of the invention.
CC   Note: The sequence data for this patent is not represented in the printed
CC   specification but is based on sequence information supplied to Derwent by
CC   the European Patent Office
XX
SQ   Sequence 1130 AA;

Query Match   5.2%;   Score 189.5;   DB 5;   Length 1130;
Best Local Similarity   21.0%;   Pred. No. 5.1e-06;
Matches 130;   Conservative 83;   Mismatches 228;   Indels 179;   Gaps 32;

RESULT 10
ABJ05434
ID   ABJ05434 standard; protein; 1138 AA.
XX
AC   ABJ05434;
XX
DT   14-NOV-2002   (first entry)
XX
DE   Human breast cancer associated polypeptide SEQ ID NO: 194.
XX
KW   Human; breast specific gene; breast specific protein; breast cancer;
KW   gene therapy; cytostatic.
XX
OS   Homo sapiens.
XX
PN   WO20026461-A1.
XX
PD   22-AUG-2002.
XX
PF   12-FEB-2002; 2002WO-US004197.
XX
PR   13-FEB-2001; 2001US-0268292P.
XX
PA   (DIAD-) DIADEXUS INC.
XX
PI   Salceda S, Macina RA, Hu P, Recipon H, Karra K, Cafferkey R;
PI   Sun Y, Liu C;
XX

```
DR   WPI; 2002-657582/70.
XX
PT   New breast specific nucleic acids and proteins, useful for identifying,
PT   diagnosing, monitoring, staging, imaging, and treating breast cancer and
PT   non-cancerous disease states in breast tissue, and in gene therapy.
XX
PS   Claim 11; Page 293-298; 367pp; English.
XX
CC   The present invention provides human breast specific coding sequences and
CC   proteins.These can be used in the diagnosis and treatment of breast
CC   cancer and non-cancerous diseases of the breast. The present sequence is
CC   a polypeptide of the invention
XX
SQ   Sequence 1138 AA;

Query Match     5.1%;    Score 186.5; DB 5;  Length 1138;
Best Local Similarity   19.3%;   Pred. No. 9e-06;
Matches 137; Conservative  90; Mismatches 235; Indels  249; Gaps   30;

Qy  125 DFAVP-------VYINSRRSQSHGRTKSGIEKEKHTPMVAPSHH------------- 162
         ||||                        :   :     :||||
Db    3 DFAAPAAAANGSSICINSSLNSSLGGAGIGVNNTPNSTPAAPSSNHPAAGGCCGSGGPGG 62

Qy  163 -SIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASS 221
                                              :: :    |:::|:
Db   63 GSAAVPKHSTVVERLRQRIEGCRRHHVNCENR----------YQQAQVEQLELER----- 107

Qy  222 HDRVNDYNASLRQDSRNRLYRDGGKTRLKDTDNGAESHLATE-NHS---------QEG 269
Db  108 RDTVSLYQRTLEQRAKKSGAGTGKQQHPSKPQQDAEAASAEQRNHTLIMLQETVKRKLEG 167

Qy  270 HGSPEDIDNDREYSKSRAC-ASLQQINEEASDDVSDDSMVDSISSI-------DVSPDDV 321
Db  168 ARSP--LNGDQQ--NGACDGNFSPTSKRIRKDIS-AGMEAINNLPSNMPLPSASP-- 217

Qy  322 VGILGQKRFWRARKAIANQQRVFAVQLFELHRL----IKVQKLIAASPDLLLDEISFLG 376
                                              ||::|  |:::|
Db  218 --------------------------LHQLDLKPSLPLQNSGTHTPG-LLEDLSKNG 247

Qy  377 KVSAKSYPVK---------KLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESS------ 421
         ::         |:|::| :
Db  248 RLPEIKLPVNGCSDLEDSFTILQSKDLKQEPLDDPTCIDTSETSLSNQNKLFSDINLNDQ 307

Qy  422 -------------------AENVVGRLSNQGHHQQS----------------NYMPFA 444
                            :::|                            :: :
Db  308 EWQELIDELANTVPEDDIQDLFNEDFEEKKEPEPFSQPATETPLSQESASVKSDPSHSPFA 367

Qy  445 N----NPPASPAPNGYCFP---------------QPPPSGNHQQWLIPVMSPSEGLIYKP 486
         :     :|::                       ::|
Db  368 HVSMGSPQARPSSSG--PPFSTVSTATSLPSVASTPAAPN-----PASSPANCAVQSP 418

Qy  487 H-PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYF-------------- 527
         | |                                      ||||
Db  419 ----PGQAPPRPGNGYLLNPAAVTVAGSASGPVAV 457

Qy  528 PPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPT 587
         ::|                    ||||||||||                  |||||
Db  458 PSSDMSPAEQLKQMAAQQQQRAKIMQQKQQ---------QQQQQQQQQQQQQQQ 510

Qy  588 KSYPRARKSRQGSTGSSPSGPQG--ISGSKSFRPFAAVDEDSNINN-APEQIMTTTTTT 644
         :|                                                 :::|
Db  511 HSNQTSNWSPLG---PSSSPYGAAFTAEKPNSPMMYPQAFNNQNPIVPPMANNLQKTTM 566

Qy  645 RTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
         :|                                             ||||
Db  567 NNYLPQNHMN------MINQQPNNLGITNSLNK------QHNILTYGNTKP 604
```

```
XX
DT   26-MAR-2002   (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 2379.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO20017042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02632.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 2379; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB7737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 771 AA;

Query Match     5.1%;    Score 186; DB 4;  Length 771;
Best Local Similarity   22.8%;   Pred. No. 5.8e-06;
Matches 109; Conservative  56; Mismatches 174; Indels 140; Gaps   22;

Qy  265 HSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI 324
         |  :             ::|:                      ::: :
Db   54 HRQGG-------ESEGPAKRAIESLVKKLKEKRDEL--DSLITAITTNGAHPSKCCVTI 102

Qy  325 LGQKRFWRARKAIANQQ---RVFAVQLF--ELHR-LIKVQKLIAASPDLLLDEISF--- 374
         :::                :::                 :::|
Db  103 --QRTLDGRLQVAGRKGFPHVIYARIWRWPDLLHKNELKHVKYCAFAFDLKCDSVCVNPY 159

Qy  375 -------------LGKVSAKSYPVKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAE 423
                      :|                                  :||
Db  160 HYERVVSPGIDLSGLSLQSGP-SRLVKDEYSAGPLVGSM------DIDGNDIGTIQHHPT 212

Qy  424 NVVGRLSNQGHHQ-QSNVMPFANNPPASPAPNGYCFP----PQPPPSGNHQQWLIP--- 474
         |:|                                            ::|
Db  213 QMVGP-GGYGYPQGPSEVVGDA-NPMSAMFPTGRTIPKIEPQDGVAGSRGSWMVPPPRL 270

Qy  475 ----------------VMSPSEGLIYK---PH--PGMAHTGHYGGYYGHYMPTPM 508
                                         ||
Db  271 GQPPQQQQQPQQQTPQPTQQQQAQSQAAAHSLPVPHGMPGMP---------GPMNPGFV 321

Qy  509 VMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQ 568
         :|                ||:          ||||||||||
Db  322 MAPPPPQQQAQNPQGNGVHHTQANSPTDPASALAMQQQQQQQQQQQQQQQQQQ------Q 371
```

```
RESULT 11
ABB58529
ID   ABB58529 standard; protein; 771 AA.
XX
AC   ABB58529;
```

```
Qy   569  QQQQRSD-----------------NEPAPQQQQQ-----PTKSYPRARK-----SR  597
          |||||                      |||                  |
Db   372  QQQQQSGGVPNGSVNAGGGAAAGGQYYGQPPVSQMQGAGGGTSVAPSVHAQQNGYVSQ  431

Qy   598  QGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVTQTTRDGG  656
            :|:||
Db   432  PGSAGSAPVGGGVFGTAQPTP----------QQPQQPPTGVQANTCSAGAQAGAGAGG  479


Db   278  -----GGNGSNTGNNTNNNGNSTN--------------NNGGSNNNGSENLTKFSVEIVQQ  319
Qy   213  IDLEKSA--SSHDRVND----YNASLRQE------SRNRLYRDGGKT  247
           ::                ::    ::  :::
Db   320  LEFTTSAANSQPQQISTNVTVKALTNTSVKSEPGVGGGGGGGGGGGGNNNNNNGGGGGG  379
Qy   248  RLKDTDNGABSHLATENHSQE------------------GHGS  272
           ::
Db   380  NGNNNNNGGDHH--QQQHQQQQQQQGGLGGLGNGNGRGGPGGGMATGPGGVAGGLGGMGM  437
Qy   273  PEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWR  332
           : :|:|    ::   ::     ::
Db   438  PPNNMSAQQKSALGNLANLVECKRPDHDFPD-----LGSLDKD-----GGGGQ  481
Qy   333  ARKALANQQRVFAVQLFELHRLIIKVQKLIAASPDLLLDE------IS  373
             :
Db   482  -----FPGFPDLLGDDNSENNDTFKDLINNLQDFNPS  513
Qy   374  FLGKVSAKSYPVKKLLPSEFL-VKPP  405
           |:|||                |:|
Db   514  FLDGFDEK--PLLDIKTEDGIKVEPNAQDLINSLNVKSEGGLGHGFGGFGLGLDNPGMK  571
Qy   406  QRGDSE------KTDQHKMESSAENVVGRLSNQGH  441
             |:|
Db   572  MRGGNPGNQQGFPNGNGTCGAPNAGGNGGNSGNLMSEHPLAAQTLKQMAEQHQHKNAM  631
Qy   442  PFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSBGLIYKPHPGMAHTGHYGGY--  499
             :|:|:                     ||:
Db   632  -----GGMGGFPRPPHGMNPQQQQQQQQQQQQQQAQQQHGQMMGGQGQPGRYND  679
Qy   500  YGHYMPTPM------------------VMPQVHPGMGFPPPGNGYFPPYGMMPTIM  537
                               :|::|                 ||
Db   680  YGGGFPNDFGLGPNGPQOQQAQQQQPQQQHLPPQFHQQKG--PGPGAGMNVQQNFLDIKQ  738
Qy   538  NPYCSS----------------QQQQQQQPNEQMNQ-----FGHPGNLQNTQQQQ  572
            ||                   ||||:                 ||||
Db   739  ELFYSSQNDFDLKRLQQQAMQQQQQQOHHQQQQQQQGGVPNFNKQQQQQQVPQQQ  798
Qy   573  -RSDNEPAPQQQQQPTKSY--------PRARKS------RQGSTGSSP----S  606
            |||||                       :                  :
Db   799  LQQQQQQQQQQQQQQQQYSPFSNQNPNAAAANFLNCPPRGGPNGNQQPGNLAQQQOQPGA  858
Qy   607  GPQGISGSKSFRPFAAVDEDSNIN--------NAPEQTMTTTTTT---RTTVTQTTRD  654
           |||                      ::|:||                  ||||
Db   859  GPQ----QQQQRGNAANGQQNNPNAGPGGNTPNAPQQQQQQSTTTTLQMKQTQLHISQQ  914
Qy   655  GGGVTRV-IKVVPH--------NAKLASENAARIFQSIQEERKRYDSSKP  695
           |||                   ||:
Db   915  GGGAHGIQVSAGQHLHLSGDMKSNVSVAAQQG--VFFSQQQAQQQQQQQP  963
```

RESULT 12
ABB63688

```
ID   ABB63688 standard; protein; 1594 AA.
XX
AC   ABB63688;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 17856.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL07791.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 17856; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1594 AA;
```

```
Query Match     4.9%;   Score 180;   DB 4;   Length 1594;
Best Local Similarity   18.2%;   Pred. No. 4.8e-05;
Matches  162;   Conservative   95;   Mismatches 306;   Indels 328;   Gaps  35;

Qy    40  EQLSIPSQRFGDHGTMNSRSNNTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFV   99
           ::  ::
Db   166  QBTTVLQKRFLE--SKNKRAAKKTDKKLPDPSQQHQQQQHQQQHQQQAQTMLA  223
Qy   100  SQMSFMENVRSSAQHDQRKMVREED------FAVPVYINSRRSQSHGRTKSGIEKEKH  152
           :|                ||
Db   224  GQL----QSSVHVQQKFLKRPAEDVDNGPDSFEPHKLPNNNNNSNSNNNNGNANANN  277
Qy   153  TPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEE  212
           :       : ::
```

RESULT 13
AAE36097

```
ID   AAE36097 standard; protein; 1133 AA.
XX
AC   AAE36097;
XX
DT   26-JUN-2003 (first entry)
XX
DE   Human Mastermind (Mam) homologue protein, MAML2.
XX
KW   Notch signal transduction; Mastermind; Mam; Mip1; Mip30; Mip6; cancer;
KW   gene therapy; cytostatic; human.
XX
OS   Homo sapiens.
XX
PN   WO200210287-A2.
XX
PD   27-DEC-2002.
XX
PF   18-JUN-2002; 2002WO-US019189.
XX
PR   18-JUN-2001; 2001US-0299112P.
```

```
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Artavanis-Tsakonas S,   Lake RJ;
XX
DR   WPI; 2003-167509/16.
DR   N-PSDB; AAD54622.
XX
PT   Inhibiting or agonizing Notch signal transduction in a cell for preparing
PT   a composition for treating or preventing cancer comprises contacting the
PT   cell with an antagonist or agonist of sumolation.
XX
PS   Disclosure; Fig 3; 84pp; English.
XX
CC   The invention relates to a method for modulating Notch signal
CC   transduction in a cell which comprises contacting the cell with an
CC   antagonist or agonist of sumolation. The invention is based on the
CC   interactions of Mastermind (Mam) protein with the Mip1, Mip30 and Mip6
CC   proteins. The method is useful for preparing a composition for treating
CC   or preventing a disease e.g. cancer. The invention is also used in gene
CC   therapy. The present sequence is human Mam homologue protein used to
CC   illustrate the method of the invention
XX
SQ   Sequence 1133 AA;

Query Match             4.9%;   Score 179.5;   DB 6;   Length 1133;
Best Local Similarity  19.0%;   Pred. No. 3.3e-05;
Matches  135; Conservative  91; Mismatches  232;   Indels  253;   Gaps   30;

Qy      125 DFAVP----------VYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHH---------------- 162
                |||||          :  :                |||||||||||||||
Db        3 DFAAPAAAANGSSICINSSLNSSLGGAGIGVNNTPNSTPAAPSSNHPAAGGCGGSGGPGG     62

Qy      163 -SIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASS    221
             |||                    |  :                 ||||: ||
Db       63 GSAAVPKHSTVVERLRQRIEGCRRHHVNCENR--------------YQQAQVEQLELER-    107

Qy      222 HDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATE-NHS--------------QEG 269
                |||                :||:          ||:|  |||:
Db      108 RDTVSLYQRTLEQRAKKSGAGTGKQQHPSKPQQDAEAASAEQRNHTLIMLQETVKRKLEG    167

Qy      270 HGSPEDIINDREYSKSRAC-ASLQQINEEASDDVSDDSMVDSISSI----------DVSPDDV 321
                |||||               ||:|    :: ||                |||||
Db      168 ARSP--LNGDQQ--NGACDGNFSPTSKRIRKDIS--AGMEAINNLPSNMPLPSASP----    217

Qy      322 VGILGQKRFWARRKAIANQQRVFAVQLFELHRL----IKVQKLIAASPDLLLDEISFLG    376
                ||||||||||||||||||||||||||||||    |||||||||||||||||:||
Db      218 ------------------------------LHQLDLKPSLPLQNSGTHTPG-LLEDLSKNG   247

Qy      377 KVSAKSYPVK----------KLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESS------- 421
                :|              ||: |           ::  ::|| :::||:
Db      248 RLPEIKLPVNGCSDLEDSFTILQSKDLKQEPLDDPTCIDTSETSLSNQNKLFSDINLNDQ   307

Qy      422 ----------------AENVVGRLSNQGHHQQS--------------------NYMPFA    444
                                :: :|    :  :::                  |||||
Db      308 EWQELIDELANTVPEDDIQDLFNEDFEKKEPEFSQPATETPLSQESASVKSDPSHSPFA    367

Qy      445 N---NPPASPAPNGYCFPP-------------QPPPSGNHQQWLIPVMSPSEGLIYKP    486
             |                               ||||: :||:  |  |||: |
Db      368 HVSMGSPQARPSSSG----PPFSTVSTATSLPSVASTPAAPN-----PASSPANCAVQSP    418

Qy      487 H-PGMAHTGHYGGYYGHYHPMMPTPMVMPQYHPGMGFPPPCNGYF------------    527
                |                         :  || ::                  :|||
Db      419 QTPNQAHT----------------------PGQAPPRPGNGYLLNPAAVTVAGSASGPVAV  457

Qy      528 PPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPT    587
                ||| :   : |||   ||||||||||:    :|     |||||||
Db      458 PSSDMSPAEQLKQMAAQQQQRAKLMQQKQQ-----------QQQQQQQ            502

Qy      588 KSYPRARKSRQGSTGSPSSPGPQG--ISGSKSFRPFAAVDEDSNINN-APEQTMTTTTTTT 644
                :  :: || :|||| : ||||:     :  ::     : :    :|
Db      503 QQQQHSNQTSNWSPLGPPSSPYGAAFTAEKPNSPMVMYPQAFNNQNPIVPPMANNLQKTTM  562
```

```
RESULT 14
ADC59316
ID   ADC59316 standard; protein; 1133 AA.
XX
AC   ADC59316;
XX
DT   18-DEC-2003  (first entry)
XX
DE   Human polypeptide #4.
XX
KW   Human; polyglutamine disease; gene; genealogical polyglutamine disease;
KW   nootropic; anticonvulsant.
XX
OS   Homo sapiens.
XX
PN   JP2002360268-A.
XX
PD   17-DEC-2002.
XX
PF   03-AUG-2001; 2001JP-00236788.
XX
PR   04-AUG-2000; 2000JP-00236839.
PR   06-APR-2001; 2001JP-00108723.
XX
PA   (KAZU-) ZH KAZUSA DNA KENKYUSHO.
PA   (DAUC ) DAIICHI PHARM CO LTD.
XX
DR   WPI; 2003-516153/49.
DR   N-PSDB; ADC59315.
XX
PT   A genealogical line diagnostic marker for polyglutamine disease, useful
PT   in the diagnosis, prevention and/or treatment, comprises a polyglutamine
PT   related gene and its encoded polypeptide.
XX
PS   Claim 4; SEQ ID NO 8; 72pp; Japanese.
XX
CC   The invention discloses polyglutamine disease related genes and their
CC   encoded polypeptides. Also claimed is a recombinant vector,
CC   transformants, preparation of the polynucleotides and resultant
CC   polypeptides, diagnostic methods and a kit. The genes and encoded
CC   polypeptides are useful in the diagnosis, prevention and treatment of
CC   genealogical polyglutamine disease. The sequence presented is a
CC   polypeptide of the invention.
XX
SQ   Sequence 1133 AA;

Query Match             4.9%;   Score 179.5;   DB 7;   Length 1133;
Best Local Similarity  19.0%;   Pred. No. 3.3e-05;
Matches  135; Conservative  91; Mismatches  232;   Indels  253;   Gaps   30;

Qy      125 DFAVP----------VYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHH---------------- 162
                |||||          :  :                |||||||||||||||
Db        3 DFAAPAAAANGSSICINSSLNSSLGGAGIGVNNTPNSTPAAPSSNHPAAGGCGGSGGPGG     62

Qy      163 -SIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASS    221
             |||                    |  :                 ||||: ||
Db       63 GSAAVPKHSTVVERLRQRIEGCRRHHVNCENR--------------YQQAQVEQLELER-    107

Qy      222 HDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATE-NHS--------------QEG 269
                |||                :||:          ||:|  |||:
Db      108 RDTVSLYQRTLEQRAKKSGAGTGKQQHPSKPQQDAEAASAEQRNHTLIMLQETVKRKLEG    167

Qy      270 HGSPEDIINDREYSKSRAC-ASLQQINEEASDDVSDDSMVDSISSI----------DVSPDDV 321
                |||||               ||:|    :: ||                |||||
Db      168 ARSP--LNGDQQ--NGACDGNFSPTSKRIRKDIS--AGMEAINNLPSNMPLPSASP----    217

Qy      322 VGILGQKRFWARRKAIANQQRVFAVQLFELHRL----IKVQKLIAASPDLLLDEISFLG    376
```

```
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 1420 AA;

Query Match             4.9%;  Score 178;  DB 4;  Length 1420;
Best Local Similarity  20.6%;  Pred. No. 6e-05;
Matches  156;  Conservative  89;  Mismatches  316;  Indels 198;  Gaps  35;

Qy      53 GTMNSRSNNTSTLVHPGPSS-----------------QPCGVERNLSVQHLDSSANQATEKFVSQM 102
              :  :       :   ::  :                 :  : ::::  ::::     : :   :
Db     557 GNLKDLESNESTVVEEETKETESDDEVIFPEPLDKTENV-------GTTANPTNE--------- 604

Qy     103 SFMENVRSAQHDQRKMVREEDFAVPVYINSRRSQSHGRTKSGIEKE---KHTPMVAPS 159
             :  :  ::  :   ::::        :   :: : ::: :::          :  ::
Db     605 ---NSVKPQAKDEVVLVSEDEDETPQNGISEKEVQT--------IQKETALKGLP--NES 652

Qy     160 SHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDV---SVTEIDL 215
              : ::  :  :  :     : :  :   :::     ::  :::::          ::
Db     653 TTESVEAKNL-QIDNSDNAC-DQFEKLKTHDMVKQTTTEDGNSNSSNLLRPAEYAEESVP 710

Qy     216 EKSASSHDRVNDYNASLRQESRNRLYR-DGGKTRLKDTDNGAESHLATENHSQEGHGSPE 274
             :::     ::  :     :  ::: :  :::  : :    :  :   :   ::   :::
Db     711 KRLRLSTDEKNEFEAETNQESLPILSRAEGG---LKDVTKRSHEHLDSSPQEEIPNKKAK 767

Qy     275 DIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI-----LGQK 328
              ::: :    : ::    ::: :      ::  : : :::  : :          :::
Db     768 TEDSDSNSSHE---GTLQIDMGGQEDKEESPKKDTQKKLDFDLNPVPEIKQNVKPLRLE 823

Qy     329 RFWRARKAIANQQRVPAVQLFLEHRLIKVQKLIA---------ASPDLLLDEI-SFLGKVS 379
              :  ::      ::    :::   : ::  :::           :   :   :  : ::
Db     824 FFKTFRRSFDTMTR-------DDLEELV-LQKVVEAMVKSDFAEIRMQLDKCESTLANYR 876

Qy     380 AKSYPVKKLLPSEFLVKPPLPHVVKQRGDSEKTDQH-----KMESSAENVVG-------- 427
             :::  ::  :: :::  : :: ::: :    :::          :  :  ::
Db     877 RKIAEVSK------QFLDLETVHKRVLK---DLEAKNSHFTAPVRITRAVGLQVGIPFKAIK 929

Qy     428 ---RLSNQGHHQQSNYMPFANNPP-ASPAPNGYCFPPQPP--------------------- 463
                : : ::  :::  :: :: :  :: :: ::  ::::
Db     930 PTVAAPESTHAAGSVLAPPSGTPPKASTSPTRATVRPRPPPTFGPSTSSSETGAGVSVT 989

Qy     464 PSGNHQQWLIPVMSPSEGLI---------------YKPHPGMAHTGHYGYGYYGHYMP 505
             ::  ::: :: ::  :::                 : :: :  : :::: ::::::
Db     990 PNSNHQQQQ-PAARSSQSLTTASVTPPVRRGCLQKVTPHRPVP---------------TNLQP 1036

Qy     506 TPMVMPQ---------------------------------PTIMN----P 539
             ::: ::                                  ::::     :
Db    1037 APMLMNQSSVQRLQSSPPIAQRTMHASKHTGTPGS--TPSSVANAINKTVMRARNSEAA 1093

Qy     540 YCSSQQQQQQPFNEQMNQFGHPGNLQNTQQQQRSDN---------EPAPQQQQQPTKSYP-- 591
             : ::: :: :: : :  : :  : : :::::              : :::::: :::
Db    1094 FLAQKQQQKQHQQQQQQQQH---LQQQQKQQQAQQQFLLRPSPGPTPGSSPAKQVPKC 1150

Qy     592 --RARKRSRQGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVT 649
               ::  :  :::: : :: : :::  : ::  : :::  ::   :   :       :
Db    1151 TTKVRAKQPPLSGATVSVPISSSSGS------ELGSYGQQQESSLAPAKPKEKAVID 1202

Qy     650 QTTRDGGGVTRVIKVV-PHNAKLASENAARIFQSIQEER 687
              ::   ::::  :    :  :  :::: ::::  ::::
Db    1203 LTDEDDAAAAAARAAQAQNARLRMEMLKRNAQTAAATR 1241


RESULT 16
ABB92874
ID   ABB92874 standard; protein; 748 AA.
XX
AC   ABB92874;
XX
DT   31-MAY-2002  (first entry)
XX
```

```
Db     218 ------------------------------LHQLDLKFSLPLQNSGTHTPG-LLEDLSKNG 247

Qy     377 KVSAKSYPVK---------KLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMRSS------- 421
             ::  :  :              :: :::  : ::  :  :   ::    ::  :
Db     248 RLPEIKLPVNGCSDLEDSFTILQSKDLKQEPLDDPTCIDTSETSLSNQNKLFSDINLNDQ 307

Qy     422 ----------------AENVVGRLSNQGHHQQS------------------NYMPFA 444
                             ::  ::::  ::  : :                  :::::
Db     308 EWQELIDELANTVPEDDIQDLFNEDFEEKKEPFSQATETPLSQESASVKSDPSHSPFA 367

Qy     445 N----NPPASPAPNGYCFPP--------------QPPPSGNHQQWLIPVMSPSEGLIYKP 486
                  :   ::   ::  ::              ::  ::  : ::   :
Db     368 HVSMGSPQARPSSSG---PPFSTVSTATSLPSVASTPRAPN------PASSPANCAVQSP 418

Qy     487 H-PGMAHTGHYGYGYYGHYMPTMVMPQYHPGMGFPPPGNGYF------------------ 527
             : ::  ::: : :::   :  : ::  :  :  ::  :::
Db     419 QTPNQAHT------------PGQAPRPGNGYLLNPAAVTVAGSASGPVAV 457

Qy     528 PPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPT 587
             ::: ::  ::::  :::::::                              :::::::
Db     458 PSSDMSPAEQLKQMAAQQQQRAKLMQQKQQ----------QQQQQQ 502

Qy     588 KSYPRARKRSRQGSTGSSPSGPQG--ISGSKSFRPFAAVDEDSNINN-APEQTMTTTTTTT 644
             ::   :: :::: : :: ::  :    : ::  : :::::::
Db     503 QQQQHSNQTSNWSPLGPPSSPYGAAFTAEKPNSPMMYPQAFNNQNPIVPPMANNLQKTTM 562

Qy     645 RTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
             :  :  :::   :::  :  :    :: :: ::::  ::::
Db     563 NNYLPQNHMN-----------MINQQPNNLGTNSLNK------QHNILTYGNTKP 600


RESULT 15
ABB63410
ID   ABB63410 standard; protein; 1420 AA.
XX
AC   ABB63410;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 17022.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
PS   Disclosure; SEQ ID NO 17022; 21pp + Sequence Listing; English.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
```

```
Db    388  PGGPEPQWGSRGPHAPHPYDYHPRGPYSSQGSYYNPSPGFGGYPPQHMPPRGGYGTDWDQRP 447
                                        ||||||
Qy    456  -YCFP---------------PQPPPSGNHQQWLIPVMSPSEGLIYKPHP------------GMAH 492
                                ::  |||:  : ::
Db    448  PYSGPYNYGRGQAQSAGPVPPSG-----PVPSPAFG-----GPPLSQVSYGYQQSH 495
                  ||:|        |||:
Qy    493  TGHYG--------GY---YGH---YMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPT 535
                  ::::          :::       ||:  || ::
Db    496  GPEYGHAAPYSQTGYQQTYGQTYEQPKYDSNPPMQPPY--GGSYPPAGGQGSYYQMQQP 553
                    ::::                |        ||:||
Qy    536  IMNPYCSSQQQQQ----------QPNEQMNQFGHP------GNLQNTQQQQQRSDNEPA 579
                  :     |||||            :                   |    |||
Db    554  GVRPYGMQQGPVQQGVGPPQPAAAASSGDVPYQGATPAAPSYGSTNMAPQQQQYTSSD 613
                          |        ||
Qy    580  PQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPPAAVDEDSNINN------- 631
                  |:    :   ::  |::   :::
Db    614  GPVQQQTYPSYSSAPPSDAYNGTQTPATGPAYQQQSVQPASSTYDQTGAQQAAAAGYGG 673
                    :::              :::
Qy    632  --APEQTMTTTTT---------TRTTVTQTTRDGGGVTR--VIKVVPHNAKLASENAAR 678
                  :::                              ::  ::
Db    674  QVAPTGGYTYPTSQPAYGSQAAYSQAAPTQTGYEQQPATQAAVYATAPGTAPVKT----- 728
                          |  ||:::
Qy    679  IFQSIQEERKRYDSSK 694
                  :|:  |::
Db    729  --QSFQSAYAQYDASQ 742
```

RESULT 17
ABB64977
ID   ABB64977 standard; protein; 2237 AA.
XX
AC   ABB64977;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 21723
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE) PE CORP NY.
XX
PI   Venter JC,   Adams M,   Li PWD,   Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL09080.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signalling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 21723; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175), and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly

---

DE   Herbicidally active polypeptide SEQ ID NO 2085.
XX
KW   Herbicidal; plant; agriculture; herbicide.
XX
OS   Arabidopsis thaliana.
XX
PN   WO200210210-A2.
XX
PD   07-FEB-2002.
XX
PF   28-AUG-2001; 2001WO-EP009892.
PR   28-AUG-2001; 2001WO-EP009892.
XX
PA   (FARB) BAYER AG.
XX
PI   Tietjen K,   Weidler M;
XX
DR   WPI; 2002-269010/31.
XX
PT   Identifying plant target proteins for herbicidally active compounds,
PT   comprising aligning and comparing nucleic acid or amino acid sequences
PT   from plant with nucleic acid or amino acid sequences from non-plant
PT   organisms.
XX
PS   Claim 5; SEQ ID NO 2085; 261pp + Sequence Listing; English.
XX
CC   The invention relates to identifying target proteins (ABB90790-ABB94016)
CC   for herbicidally active compounds, comprising aligning and comparing
CC   nucleic acid or amino acid sequences from plant with nucleic acid or
CC   amino acid sequences from non-plant organisms using suitable search
CC   parameters, where plant sequences having an E-value greater by a factor
CC   of 3 than the E-value of most similar non-plant sequences are selected.
CC   The polypeptides or nucleic acids encoding them are useful for
CC   identifying modulators. The identified modulators are useful as
CC   herbicides
XX
SQ   Sequence 748 AA;

```
Query Match      4.8%;   Score 177;   DB 5;   Length 748;
Best Local Similarity  19.5%;   Pred. No. 3e-05;
Matches  155;   Conservative 87;   Mismatches 274;   Indels 280;   Gaps 35;

Qy    123  EEDFAV-----PVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHS------IRFQEVNQ 171
                  |||| |          :                    ::
Db      3  EEDFVVISTPPVDLKRKLD-----------EVELNGNIVDDSNQTSDSSQAKRAKLEDEAQ 52
Qy    172  TGSKQNVCLAT------CSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHD 223
                  :  ::            ::                :  :
Db     53  DGLGVSVSLASYKLNLSHCCSSAFVVDY-GNTQENG--SSMEVKEEQLQEPKEE---- 104
Qy    224  RVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDI------ 276
                  :         :    :                :          :  :
Db    105  -NQDSVPLVEBVQDPIHADESENKICSVDQPTDDQVKPEDNQQ--GCIEDVNGGEHQK 159
Qy    277  --DNDREYSKSRACASLQQINEEASDDVSD--DSMVDSISSIDVSPPDVVGILGQKRF 330
                  ::  : :        :  :        ||||        ||:
Db    160  VIDDDSKEVNDGSSQENGKEENKEVDGEKSQKEVDGTQSTTRRIDV-PSSKVGVL---- 213
Qy    331  WRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVS--------- 379
                  : : :                           :  :
Db    214  -----IGKGGETIRYLQFNSGAKIQILRDSEADPSSALRPVEIIGSVACIESAEKLISA 267
Qy    380  --------AKSYPVKKL--LPSEFLVKPLPHV--VVKQRGDSEKT-------- 413
                  : ::        :  :         ::
Db    268  VIAEAEAGGSPALVARGHPSTHAIGIPEQIEIKVPNDKVGLIIGRGGETIKNMQTRSGAR 327
Qy    414  ---DQH----------KME-SSAENVVGRLSNQ------GHHQQSNYMP------- 442
                   :               : ::                  |
Db    328  TQLIPQHAEBGDGLKERTVRISGDKMQIDIATDMKDVMNQNARPSSYSGGYNQPAYRPQG 387
Qy    443  ----------------FANNPPASPAPNG---------- 455
                                    || ||
```

```
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 2237 AA;

Query Match             4.8%;  Score 176;  DB 4;  Length 2237;
Best Local Similarity  20.8%;  Pred. No. 0.00016;
Matches  171;  Conservative  92;  Mismatches  274;  Indels  284;  Gaps  41;

Qy      7 EEKILEP--------MFPRLHVNDADKGG------PRAPPRN----------KMALYRQL 42
            |:|:||        ::|:||            ||:||
Db   1284 EQLIVEPPQTQPLELLPRASSANSDYSGYRIRPPSRPSSNSSQPDYTQVSPAKMALRRHL 1343

Qy     43 SIPSQRFGDHGTMNSRSNNTSTLVHPGFSSQPCG----------VERNLSVQHLDSSAA 91
Db   1344 S--QEKLSQHVTPQ------ATPPLPGHGGAPTSGKTIGDLVNGEIERTLEISH--QSII 1393

Qy     92 NQATEKFVSQMSFME----NVRS----SAQHDQRKMVR---EEEDFAVPVYINSRR 136
Db   1394 NAAVNMSTSGASFMERAFLNERSNDRLLINLNAQRPERVHVRPLSEESQDPQPTSYAQER 1453

Qy    137 SQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANA 196
Db   1454 GPGLG---AGGAAAGGNSNlATLAH--VAYAQKAQGGARAN--AGTAPPATH---SSSA 1502

Qy    197 RSG-----------GFVISLDVSVTEEIDLRKSASSHD---------RVNDYN------ 229
Db   1503 RSGRDYQPVALPRAELKG---SIRAYFHEBQQQKQSKGAGSAGSSSLRGPRLNGANPPLE 1559

Qy    230 --ASLRQESRNRLYRDGGKTRLKDTDNGAES----------------HLATENHSQEGHG 271
Db   1560 GLAASLQDHVRARKYKEETEERQRRAAAAASSSAGPPAGMELPTHYAHQAPPAHSYHHHG 1619

Qy    272 SPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFW 331
Db   1620 -------------------------------ASINGTPHKVB--LGIKR- 1635

Qy    332 RARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPS 391
Db   1636 --SSPLAPHQQP--------PRPSKLAHYEPPTTQQQ----HAHAHLYANGQVLP- 1676

Qy    392 EFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQCHHQQSNYMPFANNPP--A 449
Db   1677 ---PPPAH--------DATTPSPTPSSSSSSCGRRSN------SNNGKLLVDPPLLM 1716

Qy    450 SPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEG--------LIYKPHPGMAHTGHYGGYYGH- 502
Db   1717 SPEINSLLGDERPLQLSHHQQQQQQMLHHHQSQQQQHLQLTQQQLRVAHLGH-GLSHGHS 1775

Qy    503 YMP--------------------------TPMVMPQYHPGMGFPPPGNGYFPPYGMMPTI 536
Db   1776 TMPTLGGQRNGNGNAADDGKLYVDQAKNSTRVSVPKYAPQQA-PPPAHVAFP--------- 1826

Qy    537 MNPYCSSQQQQQQQQPNEQMNQFGHPGNLQNTOQQQRSDNEPAPQQQQQPTKSYPRARKS 596
Db   1827 LEPHILHQHKYQRLATQTVAED------LTMKRKEQESDSELEP--EVVAAFVEDGGS 1876

Qy    597 RQGSTGSSPSSGPPQGISGSKSFRPFAAVDE--DSNINNAPEQTMTTTT--------- 642
Db   1877 HSGSSSSSSSSH--SGSDSNDVIAADDETHWQHRVSSGFDRLVAFASTELDKTRRSIDG 1933

Qy    643 -TTRTTVTQTTRDGGVTRVIKVVPHNAKLASENAARIFQS 682
Db   1934 DTVPASISCNTSPDSGIT-------HSS--SNSDAVRTFLS 1965

RESULT 18
AAG85017
ID   AAG85017 standard; protein; 757 AA.
XX
AC   AAG85017;
XX
DT   06-AUG-2003 (revised)
DT   11-SEP-2001 (first entry)
```

```
XX
DE   Shrimp white spot Bacilliform virus (WSBV) protein 108.
XX
KW   Shrimp white spot Bacilliform virus; WSBV; diagnosis; viral infection;
KW   antiviral agent; gene expression; antisense construct;
KW   transgenic viral resistant shrimp.
XX
OS   White spot syndrome virus.
XX
PN   WO200138351-A2.
XX
PD   31-MAY-2001.
XX
PF   08-NOV-2000; 2000WO-US028888.
XX
PR   24-NOV-1999;    99CN-00124717.
XX
PA   (PENY-) PE CORP NY.
PA   (THIR-) THIRD INST OCEANOGRAPHY STATE OCEANI C A.
PA   (SINO-) SINOGENOMAX CO LTD.
XX
PI   Xu X,  Yang F,  He J,  Pham L,  He M,  Ye Y,  Shen Y,  Kodira C;
XX
DR   WPI; 2001-355877/37.
DR   N-PSDB; AAH62797.
XX
PT   primary nucleotide sequence of the shrimp white spot Bacilliform virus
PT   (WSBV), useful for producing viral polypeptides that can be used to
PT   screen for agents that are useful for treating WSBV infection.
XX
PS   Claim 1; Fig 3; 626pp; English.
XX
CC   The invention provides the primary nucleotide sequence of the WSBV genome
CC   (AAH62689), predicted transcript sequences (AAH62689-AAH62839) and
CC   encoded proteins (AAG84910-AAG85051) and oligonucleotide sequences
CC   (AAH62840-63160) suitable for use as primers or probes. The nucleic acid
CC   molecules and proteins of the invention are useful for diagnosis and
CC   monitoring viral infection, in screens for antiviral agents and for
CC   monitoring viral gene expression or activity during a treatment regimen.
CC   The nucleic acid molecules are also useful as antisense constructs to
CC   control viral gene expression in infected cells and tissues and to create
CC   transgenic viral resistant shrimp. (Updated on 06-AUG-2003 to correct OS
CC   field).
XX
SQ   Sequence 757 AA;

Query Match             4.8%;  Score 174.5;  DB 4;  Length 757;
Best Local Similarity  21.3%;  Pred. No. 4.9e-05;
Matches  155;  Conservative  80;  Mismatches  274;  Indels  219;  Gaps  33;

Qy      2 KRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLS1PSQRFGDHGTMNSRSNN 61
            |||:|         ||||
Db    123 KEEKDGEEEBEENVEDEHVTPTTSVSKRAKQMKK-----HIFPPSKK-----RKRSDT 170

Qy     62 TSTLVHPGSSQPCGVERNL----------SVQHLDSSAANQ-----ATEKFVSQMSFMEN 107
Db    171 ESKAV---PAGKMMTVSRPLRGAITSGSILGVRSENAPQYDYVSYLADEAVVKEKAIQYR 227

Qy    108 VRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQ 167
Db    228 IRSLLAN----LLKANKTKAFPT-SSSLLSSEQGKKKFG-GKRTNTFVVTNVGAELVK-- 279

Qy    168 EVNQTGSKQNVCLATCSKPEVRD-QVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVN 226
Db    280 ----ALLANSCWAISHRKDIRSGEIQWQELSSKILKSLNDGNATEINNLMSSIVEDRI- 333

Qy    227 DYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSR 286
Db    334 ----QRTVKERVYFE-----------------QLATVCNNLFGTRILPNKNFDKNFV--- 369

Qy    287 ACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAV 346
Db    370 ----------SVASDNSNATVRGLSI-------------PRYFR--AINNNVWVKMS 401
```

```
CC     sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC     ABB72072). The sequence data for this patent did not form part of the
CC     printed specification, but was obtained in electronic format directly
XX     from WIPO at ftp.wipo.int/pub/published_pct_sequences
SQ     Sequence 1187 AA;

Query Match            4.7%;    Score 172;  DB 4;  Length 1187;
Best Local Similarity  20.1%;   Pred. No. 0.00014;
Matches 155;  Conservative 106;  Mismatches 293;  Indels 218;  Gaps 35;

Qy  51 DHGTMNSRSNNTSTIVHPGPSSQPCGVERNLSV-QHLDSSAANQATEKFVSQMSFMENVR 109
        :||:||      :  : : : .                EHELIVDDELPDPANGASDDLGLQLQVGEQVT
Db  82 EDGTLVTESRRTT----------------------EHELIVDDELPDPANGASDDLGLQLQVGEQVT 127

Qy 110 SSAQHDQRKMVREEDFAVPVYINSRRSQSHGR---TKSGIEKEHTPMVAPSSHHSI-- 164
Db 128 TTAAN--QRFFRQRDEQHVDVLANGRLQSTEMRYAAETTQMDKDGPTDLERPGDWDSLSD 185

Qy 165 RFQEVNQTGSKQ-NVCL---ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLE 216
Db 186 RMRKLRRSQQQQKEVALLADRKDALTKRPLDFDR-EEETRKGETMKWLE---SHFGSE 240

Qy 217 KSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGS--- 272
Db 241 STASNDSRDDEADAEVEPTKKSYF----NVTIK-SQSHAQSHLATQQHSPQHISHGHH 294

Qy 273 ----PEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISIDVSPDDVVGILGQK 328
Db 295 PQTLPRSAEREREREREREREREREREREREDRDRHPLAMTASASQERKAPANSAG--GRS 352

Qy 329 RF------WRARKAIANQQRVFAVQLFELHRLIKVQKLTAASPDLLLDEISFLG---KVS 379
Db 353 KYFQGISNWSERKDPGPPVNHFSSQAFR----------DDLQETIRRNGLKKKVS 397

Qy 380 AKSYPVKKLLPSEFLVKPLPHV----VVKQ---------RGDSEKTDQ-- 415
Db 398 GPGSSQGSALAQETVVQPKDSYVSREDIIQQKRQSLSSQFLARSMRGSRDALDDLEAIPG 457

Qy 416 ------HKMESSAENVV-GRLSNQGHHQQSNYMPFANNPASPAPNG---YCFPPQ 461
Db 458 PGPGPSMISPRAEEPSHKRVAYGHRGNNGTHINGH---ALNGVAPTLVNGRGKSYEPPA 513

Qy 462 PPPSGNHQQWLIPVMSPSEG---LIYKPHPGM------------AHTGHYGGYY 500
Db 514 PPPAA-------VPPLEFGGGSQLGQNPRPTVPQRRRAIEKKLGEHSHTHSHSHSHSHHT 567

Qy 501 GH--------YMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQPN 552
Db 568 SHSAHLPVKSQAPMSQLNQLEPPPDYSPPPR----------SRSRSRSSPQVD 610

Qy 553 EQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTG-- 602
Db 611 YLMTREMPASTMTLSRKQQQRT--------RFAPTSSYPPA--TGGTGNLRPPGATSV 658

Qy 603 ----SSPSGPQGISGSKSFRPFAAVDED----SNINNAPEQ--TMTTTTTRTTVT 649
Db 659 STSNLSAMSG--GVSTSKSSSRMGQVIGNSLRKLVSKIRSASAERKFRMKSAASKSREQSP 716

Qy 650 QTTRDGGGVTRVI-----------KVVPHNAKLA-SENAARIFQSIQEERKR 689
Db 717 SPGQDPRGVGPPVTTYQQYNVIDGHGSHDSEDNSDNQRRERERERERER 768
```

```
Qy 347 QLFEL------------HRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLIV 395
Db 402 STMDLLVGGGMRSKSEHSISMLEKCAA----------GVLARASAR--PVEKMIKSAV-- 447

Qy 396 KPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNG 455
Db 448 --------------EETSQAFNLSTGVFVPKQQQQQRQQQQQQFPPF-QPPPFPLP-- 488

Qy 456 YCFPPQQPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVM--PQY 513
Db 489 --PPQAP----------FQVVQPTYQGVLNPY--------YQYNQYNPYAPQQLQQQYPLY 530

Qy 514 HPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQR 573
Db 531 FLGNQSQPP--------PQL---QQQQQQQPPQPNNIPPP---PTPQQQSP 569

Qy 574 SDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAP 633
Db 570 SNIPPPQQQQQP---FP------VQLISSPPPPIP----------NTAP 602

Qy 634 EQTMTTTTTTTRTTVTQTTRDGGGVTRVIK-----VVPHNAKLASENAARI-FQSIQEER 687
Db 603 SPPISRVRFDSRSTTPQPP----PTPVLPKPTLPLPPSTARAEEENATDMSFTDIDSEL 657

Qy 688 KRYDSSKP 695
Db 658 GSIDFDLP 665
```

```
RESULT 19
ABB67666
ID   ABB67666 standard; protein; 1187 AA.
XX
AC   ABB67666;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 29790.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL11769.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 29790; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
```

```
RESULT 20
ABB58382
ID   ABB58382 standard; protein; 3502 AA.
XX
AC   ABB58382;
XX
DT   26-MAR-2002 (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 1938.
```

```
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-0614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li FWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL02485.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 1938; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 3502 AA;

Query Match             4.7%;  Score 172;  DB 4;  Length 3502;
Best Local Similarity   17.3%;  Pred. No. 0.00064;
Matches  121;  Conservative  76;  Mismatches  226;  Indels  278;  Gaps  25;

Qy      26 GGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSNNTSTLVHPGPSSQPCGVERNLSVQH 85
                 ||| |       :        ||     : :     :: ::
Db      13 GAP---PPTNN------NNSSSSSSSNSNNNNTSNNNNNSNSGNNNTPNGAQQQQQQHH 63

Qy      86 LDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEDFAVPVYINSRRSQSHGRTKS 145
                                                   ||| :      : ::::::
Db      64 QQQQQHHQQ---------QRVVASSNQQQQQQLQQQ---------------QQQQQQ 100

Qy     146 GIEKEKHTPMVAPSSHH------SIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSG 199
               ::::|  ||||
Db     101 QQQQQQHDN--GPTSHHLHQSSISISNQQNKATQQQ----------------- 135

Qy     200 GFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESH 259
           ------                    :  ::           :  ::
Db     136 ------------QQQQQHQKQQQQQASSMSAYQQRL------PSNAASIH 168

Qy     260 LATRNHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPD 319
                                                                :
Db     169 ------RRPHWSYRALEQQQQYAKQ--AAHLQQQQHQSHQ---------------- 201

Qy     320 DVVGLIGQKPFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVS 379
                                                             :
Db     202 --------------QQQQQQQDQDRTNLHLQIH--------------------- 220

Qy     380 AKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHH--QQ 437
                :  :        :    : :::   |  :   |  :   :      : : :
Db     221 -------------HHHQQQQQQQQQQQQQQQQQQQQ---QQKQQQHMQQQ 256
```

```
RESULT 21
AAB36095
ID   AAB36095 standard; protein; 1596 AA.
XX
AC   AAB36095;
XX
DT   26-JUN-2003  (first entry)
XX
DE   Drosophila sp. Mastermind (Mam) protein.
XX
KW   Notch signal transduction; Mastermind; Mam; Mip1; Mip30; Mip6; cancer;
KW   gene therapy; fruitly; cytostatic.
XX
OS   Drosophila sp.
XX
PN   WO2002102987-A2.
XX
PD   27-DEC-2002.
XX
PF   18-JUN-2002; 2002WO-US019189.
XX
PR   18-JUN-2001; 2001US-0299112P.
XX
PA   (GEHO ) GEN HOSPITAL CORP.
XX
PI   Artavanis-Tsakonas S, Lake RJ;
XX
DR   WPI; 2003-167509/16.
DR   N-PSDB; AAD54620.
XX
PT   Inhibiting or agonizing Notch signal transduction in a cell for preparing
PT   a composition for treating or preventing cancer comprises contacting the
PT   cell with an antagonist or agonist of sumolation.
XX
PS   Disclosure; Fig 1; 84pp; English.
XX
CC   The invention relates to a method for modulating Notch signal
CC   transduction in a cell which comprises contacting the cell with an
CC   antagonist or agonist of sumolation. The invention is based on the
CC   interactions of Mastermind (Mam) protein with the Mip1, Mip30 and Mip6
CC   proteins. The method is useful for preparing a composition for treating
CC   or preventing a disease e.g. cancer. The invention is also used in gene
CC   therapy. The present sequence is Drosophila sp. Mam protein used to
CC   illustrate the method of the invention
XX
SQ   Sequence 1596 AA;

Query Match             4.7%;  Score 171.5;  DB 6;  Length 1596;
Best Local Similarity   18.1%;  Pred. No. 0.00024;
Matches  161;  Conservative  93;  Mismatches 311;  Indels  325;  Gaps  35;

Qy      40 EQLSIPSQRFGDHGTMNSRSNNTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFV 99
```

```
Qy     438 SNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYG 497
                                                  :   :||
Db     257 QQQQPLS----------------PPHPPGSSSNSSSAAAAAAAAAAAAAAAAVNPG----- 294

Qy     498 GYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGM----MPTIM-NPYCSSQQQQQQPN 552
                :  |                    |:    :: ||||    :|       ||||
Db     295 ----YPPSSAAAAVNSGYPPRPPQHRFIQNTGYSIAPAPTYRDNPYSRHTQIQQQQQQ 349

Qy     553 EQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKS--------------- 596
               :|:                  :  ||
Db     350 QQQQQ---------QQQQQQQ---QQQQQAAASMPFYQRAAARARAVAAVSAGKGNV 393

Qy     597 --RQSTGSPSPGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRD 654
               :  |              |                :||  ::|| :||||||||
Db     394 SGQSNSNSSSSSGGGGGGSAG-------GSAP---------PGGGVQVSQS 430

Qy     655 GGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
                                        |  ::|  :||
Db     431 GG--VLVMEAMPHYASQPNSNPS---QQQQQHQQQQGGNP 466
```

```
KW   nervous system disorder; arthritis; inflammation.
XX
OS   Homo sapiens.
XX
PN   WO200157190-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US004098.
XX
PR   03-FEB-2000; 2000US-00496914.
PR   27-APR-2000; 2000US-00560875.
PR   20-JUN-2000; 2000US-00599805.
PR   19-JUL-2000; 2000US-00620325.
PR   01-SEP-2000; 2000US-00654936.
PR   15-SEP-2000; 2000US-00663561.
PR   20-OCT-2000; 2000US-00693325.
PR   30-NOV-2000; 2000US-00728422.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y;
PI   Ma Y, Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
PI   Xue AJ, Yang Y, Wejhrman T, Goodrich R;
XX
DR   WPI; 2001-476283/51.
DR   N-PSDB; AAK53100.
XX
PT   Nucleic acids encoding polypeptides with cytokine-like activities, useful
PT   in diagnosis and gene therapy.
XX
PS   Claim 20; Page 396-397; 6221pp; English.
XX
CC   The invention relates to polynucleotides (AAK51456-AAK53435) and the
CC   encoded polypeptides (AAM78823-AAM80302) that exhibit activity elating to
CC   cytokine, cell proliferation or cell differentiation or which may induce
CC   production of other cytokines in other cell populations. The
CC   polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC   peptide therapy. The polypeptides have various cytokine-like activities,
CC   e.g. stem cell growth factor activity, haematopoiesis regulating
CC   activity, tissue growth factor activity, immunomodulatory activity and
CC   activin/inhibin activity and may be useful in the diagnosis and/or
CC   treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC   inflammation. Note: Records for SEQ ID NO 2110 (AAK52581), 2111
CC   (AAK52582) and 3666 (AAM80020) are omitted as the relevant pages from the
CC   sequence listing were missing at the time of publication
XX
SQ   Sequence 843 AA;

Query Match             4.6%;   Score 170.5; DB 4; Length 843;
Best Local Similarity   23.8%;  Pred. No. 0.00012;
Matches   86;  Conservative   34;  Mismatches  144;  Indels  97;  Gaps  15;
```

```
Qy  397 PPLPHVVKQRGDSEK------TDQHKMESSA---ENVVGRLSNQGHHQQSNYMPFANNPP 448
            :   ::  :|:|:|       ||:||| :       ::   ||||
Db   32 PKKREIPATSRSSEEKAPTLPSDNHRVEGTAWLPGNPGGRGHGGGRH-------------GPA 81

Qy  449 ASPAPNGY-------------------------------CFPQPPPSGNHQQWLIPVMSPSEGLIYKPHPG----- 489
                                                  :|
Db   82 GTSVELGLQQGIGLHKALSTGLDYSPPSAPRS------------------VPVATTLPAAYATPQPGTPVS 134

Qy  490 ---MAHTGH-------------YGGYGYHMPTPMVMPQYHPGMGFPPPGNGYFPP----- 529
            :         :        ::|::|: :
Db  135 PVQYAHLPHTFQFIGSSQYSGTYASFIPSQLIPPTANPVTSAVASAAGATTPSQRSQLEA 194

Qy  530 YGMMPTIM----------NPYCSSQQQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQ 584
                               :: |:|::     ||
Db  195 YSTLLANMGSLSQTPGHKAEQQQQQQQQQQQ------QQQQQQQQQQHQQQQQQQQQQ 247

Qy  585 QPTKSYPRARKSRQGSTGSSSPSGPQG----ISGSKSFRPFAAVDEDSNINNAPEQTMT-- 638
             :                            :: ||||
Db  248 QQQQQQHLSRAPGLITPGSPPPAQONQYVHISSSPQNTGRTASPAIPVHLHPQTMIPH 307
```

```
Db  166 :: :: :||: :|: : |||          |||
        QETTVLQKRFLE--SKNKRAAKTDKKLLPDPSQQHQQQQHQQQHQHQQAQTMLA 223
Qy  100 SQMSFMENVRSSAQHDQRKMVREEED-------FAVPVYINSRRSQSHGRTKSGIEKEKH 152
        :|:          |:
Db  224 GQL------QSSVHVQQKFLKRPAEDVDNGPDSFEPPHKLPNNNNNSNSNNNGNANANN 277
Qy  153 TPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTBE 212
                                              |:||
Db  278 ------GGNGSNTGNNTNNNGNSTN-------NNGGSNNNGSENLTKFSVEIVQQ 319
Qy  213 IDLEKSA--SSHDRVND------YNASLRQE------SRNRLYRDGGKT 247
        ::          |:|:|
Db  320 LEFTTSPANSQPQQISTNVTVKALTNTSVKSRPGVGGGGGGGGNSGNNNNNGGGGGGG 379
Qy  248 RLKDTDNGAESHLATENHSQE-------------------GH 270
        |:
Db  380 NGNNNNNGGDHHQQQQHQHQQQQQGGGLGLGNNGRGGGPGGMATGPGGVAGGLGGM 439
Qy  271 GSPEDINDREYSKSRACASLQQINEREASDDVSDDSMVDSISIDVSPDDVVGILGQKRF 330
                                              ::|: |:
Db  440 GMPPNMMSAQQKSALGNLANLVECKREPDHDFPD-----LGSLDKD----GGGGQ-- 485
Qy  331 WRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLDE------------- 371
                                                      |||:
Db  486 ---------------------------FPGFPDLLGDDNSENDTFKDLINNLQDFN 515
Qy  372 ISFLGKVSAKSYPVKKLLPSEFL-VKPP------------------LPHVV 403
                                                        |:
Db  516 PSFLDGFDEK--PLLDIKTEDGIKVEPPNAQDLINSLNVKSEGGLGHGFGGFGLGLDNPG 573
Qy  404 VKQRGDSE-------KTDQHKMESSAENVVGRLSNQGH------------HQQSN 439
        :|:|:        ||
Db  574 MKMRGGNPGNQGGFPNGPNGGTGGAPNAGGNGGNSGNLMSEHPLAAQTLKQMAEQHQHKN 633
Qy  440 YMPFANNPPASPAPNGYCFPPQPPPSSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGY 499
                        ||                                      |:|
Db  634 AM-----------GGMGGFPRPPHGMNPQQQQQQGQQQQAQQGHGQMMGGQGPGRY 681
Qy  500 --YGHYMPTPM-----------------VMPQYHPGMGFPPPGNGYFPPYGMMP 534
           |:                     :
Db  682 NDYGGGFPNDFGLGPNGPQQQQQQAQQQPQQQHLPPQFHQQKG-PGPGAGMVQQNFLD 740
Qy  535 TIMNPYCSSQ-------------------QQQQQPNEQMNQ----FGHPG-NLQNTQQ----QQQR 573
           |:|                       ||||
Db  741 IKQELFYSSQNDFDLKRLQQQQAMQQQQQQHQQQQPKMGGVPNFNKQQQQQVPQQQL 800
Qy  574 SDNEPAPQQQQQPTKSY--------PRARKS------RQGSTSSSP-------SGP 608
           :|                                 |:
Db  801 QQQQQQQQQQQQYSPFSNQPNAAANFLNCPPRGGPNGNQQPGNLAQQQQQPGAGP 860
Qy  609 QGISGSKSFRPFAAVDEDSNIN------NAPEQTMTTTTTTTTRTTVTQT----TRDG 655
        |:                                  :
Db  861 Q----QQQRGNAGNGQQNNPNTGPGGNTPNAPQQQQQQQSTTTTLQMKQTQQLHISQQG 916
Qy  656 GGVTRV-IKVVPH------------NAKLASENAARIFQSIQEERKRYDSSKP 695
        ||                                      ::
Db  917 GGAQGIQVSAGQHLHLSGDMKSNVSVRAQQG--VFFSCQQAQQQQQQQP 964
```

```
RESULT 22
AAM79967
ID   AAM79967 standard; protein; 843 AA.
XX
AC   AAM79967;
XX
DT   06-NOV-2001 (first entry)
XX
DE   Human protein SEQ ID NO 3613.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW   vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW   tissue growth factor; immunomodulatory; cancer; leukaemia;
```

```
Qy    639  TTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQ-------EERKR 689
                |:|:||:|                    ||:||::||::|||| |   :| :|
Db    308  TLTLGPSQVVMQYADSGS-----HFVPREATKKAE-SSRLQQAIQAKEVLNGEMEKSRR 361

Qy    690  Y 690
             |
Db    362  Y 362


RESULT 23
ABB71657
ID    ABB71657 standard; protein; 2328 AA.
XX
AC    ABB71657;
XX
DT    26-MAR-2002  (first entry)
XX
DE    Drosophila melanogaster polypeptide SEQ ID NO 41763.
XX
KW    Drosophila; developmental biology; cell signalling; insecticide;
KW    pharmaceutical.
XX
OS    Drosophila melanogaster.
XX
PN    WO200171042-A2.
XX
PD    27-SEP-2001.
XX
PF    23-MAR-2001; 2001WO-US009231.
XX
PR    23-MAR-2000; 2000US-0191637P.
PR    11-JUL-2000; 2000US-0614150.
XX
PA    (PEKE ) PE CORP NY.
XX
PI    Venter JC, Adams M, Li PWD, Myers EW;
XX
DR    WPI; 2001-656860/75.
DR    N-PSDB; ABL15760.
XX
PT    New isolated nucleic acid detection reagent for detecting 1000 or more
PT    genes from Drosophila and for elucidating cell signaling and cell-cell
PT    interactions.
XX
PS    Disclosure; SEQ ID NO 41763; 21pp + Sequence Listing; English.
XX
CC    The invention relates to an isolated nucleic acid detection reagent
CC    capable of detecting 1000 or more genes from Drosophila. The invention is
CC    useful in developmental biology and in elucidating cell signalling and
CC    cell-cell interactions in higher eukaryotes for the development of
CC    insecticides, therapeutics and pharmaceutical drugs. The invention
CC    discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC    sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC    ABB72072). The sequence data for this patent did not form part of the
CC    printed specification, but was obtained in electronic format directly
CC    from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ    Sequence 2328 AA;

Query Match          4.6%;  Score 170;  DB 4;  Length 2328;
Best Local Similarity  19.4%;  Pred. No. 0.00053;
Matches  122;  Conservative  79;  Mismatches  219;  Indels  208;  Gaps  29;

Qy    152  HTPMVAPSSHHSIRFQE--VNQTGSKQNVCLATCSK--PEVRD--QVKANARSGGFVISL 205
                |::| ::||: ||:| ||:| | ||: :  ::|  | | |    :|:|| :| ||
Db   1740  HSTLVSDDERHYMNLMKKLKKEIGEKNNASIRVIRQLLPLYKQPTEVIACEHSG-----M 1794

Qy    206  DVSVTEEIDLEKS---ASSHDRVNDYN----------------ASLRQESNRLYRDGGK 246
            |::| :::   ::       |                        ||:|| : ||
Db   1795  DTKGNKICDIDKKQLRISDKQRISVWDILEGHKNPAPLSWVWFGAVKLERKPLTYEB--- 1851

Qy    247  TRLKDTDNGAESHLATENHSQEGHGS--------PEDIDNDREYSKSRACASLQQINEE 297
                     :  |    |  |:||:        |||:  | |: | | : :
```

```
Qy   1852  -------AHRNLKYHTHSLVKPSSYYYEPLPLPPEDIEPVPE---KIC--------- 1889
                                                    | |
Db    298  ASDDVSDDSMVDSLSSIDVSPDDVVGILGQKR---FWRARK------AIANQQRVFAV 346

Qy   1890  ----IKDEMKADTPSSVDQSPSAVVGTGRGRGKGTTTRKRKPKNPKTPVVNTQ----- 1940
                                                    :  :: |
Db    347  QLFELHRLIKVQKLIAASP--------DLLLDEISFLGKVSAKSYPVKKLLPSEFIVKP- 397

Qy   1941  -------QQQPQLAQQPQQPQNVQQQLQQQQQQHMQQQMHQQQOMPNQMQMGMQOMPM 1991
                     ::|:|:|||:|:  |::|::| || :::: ::
Db    398  -------PLPHVVVKQRGDSRKTDQHKMESSAENVVGRLSNQG-----HHQQSNYMP 442

Qy   1992  NMPMNMQQFAPNPNMNMQQNAMLQQQQQQMQQQMGNNPMQQQLNVGGNGQPNPQMNFM- 2050
               ::|: :::                     :  |:
Db    443  FANNPPASPAPNGYCFPPQPPSSGNHQQWLLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGH 502

Qy   2051  ------QQGPGGGGAGPQGMP-GQQQQW------------------------H 2072
                                  :::
Db    503  YMPTPMVMPQ-YHPGMGFPPPGNGYFPPYGMMPT--IMNP------YCSSQQQQQQQPN 552

Qy   2073  NAPQQQQPPQPYH--------NQYAPHQQNMQSNRIERPPLNANSKQALSQMLRQRPF 2123
                                    ::|::| :
Db    553  EQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQQTKSYPRARKSRQGSTGSSPSGPQGIS 612

Qy   2124  QQQAQQGPGGGFNPMQQQPQASQQQPGPQQQMNPNQMRQQQMNPQQ-----NPQSVA 2175
                                         :|::
Db    613  GSKSFRPFAAVDEDSNINNAPEQTMTTTTTTR---------------TTVTQTTRDGGGV 658

Qy   2176  A------FNAMQQQPQ-QNAQQQQMNPNQQQQQQFMRGGNMRPGMAPNQMNQMNQMGGQGM 2228
                                                    :  : |  :|
Db    659  TR---VIKVVPHN-AKLASENAARIFQS 682
                   : :|:  :| ::: ::
Db   2229  SQNPMMQQQITPQNMVGMVNPNANQMMQS 2256
```

```
RESULT 24
ABB62536
ID    ABB62536 standard; protein; 333 AA.
XX
AC    ABB62536;
XX
DT    26-MAR-2002  (first entry)
XX
DE    Drosophila melanogaster polypeptide SEQ ID NO 14400.
XX
KW    Drosophila; developmental biology; cell signalling; insecticide;
KW    pharmaceutical.
XX
OS    Drosophila melanogaster.
XX
PN    WO200171042-A2.
XX
PD    27-SEP-2001.
XX
PF    23-MAR-2001; 2001WO-US009231.
XX
PR    23-MAR-2000; 2000US-0191637P.
PR    11-JUL-2000; 2000US-0614150.
XX
PA    (PEKE ) PE CORP NY.
XX
PI    Venter JC, Adams M, Li PWD, Myers EW;
XX
DR    WPI; 2001-656860/75.
DR    N-PSDB; ABL06639.
XX
PT    New isolated nucleic acid detection reagent for detecting 1000 or more
PT    genes from Drosophila and for elucidating cell signaling and cell-cell
PT    interactions.
XX
PS    Disclosure; SEQ ID NO 14400; 21pp + Sequence Listing; English.
XX
CC    The invention relates to an isolated nucleic acid detection reagent
```

```
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL1176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB5737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 333 AA;

Query Match            4.6%;  Score 168.5; DB 4;  Length 333;
Best Local Similarity 28.8%;  Pred. No. 4.8e-05;
Matches  69;  Conservative  20;  Mismatches  78;  Indels  73;  Gaps  14;

Qy   429 LSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQQPPSGNH-----QQWLIPVMSPSEGL 482
         ::|:|  ||| :   |  |||     ::       :  | ||
Db     8 MNSQGPHDQQH--PQAGN--NMYPGGN-FQQQQPPGNHFADPRGPNFFIPDMS----- 56

Qy   483 IIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPP---PGNGYFPPYGMMPTIMNP 539
         |  ||||   |   ||| |    :|  :   |   |   |    |||:|  | || :
Db    57 --KPPPG------FPGPMHHNQGPAMGQQQPGQ-FPPMQAPGSGNEPPVDLGVLSAAI 107

Qy   540 YCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQ---PTKSYPR--ARK 595
         ||: || |   :|   :|||  :|     :|||| || ||||   :|  | |  |||
Db   108 QACLQMQQQHHPDDPQQQGQH--QLQQQQQQQ------PQQQKQPPPMAAYPQNIVRP 157

Qy   596 SRQSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTTQTTRDG 655
           : :  :|:  :::::::|  |||| |::: |   ||  |||||||||: : || |:|
Db   158 GQQNADSS---------------------EVDLDVLNA--------AIRAVITNKPEDG 187

RESULT 25
ABB71899
ID   ABB71899 standard; protein; 506 AA.
XX
AC   ABB71899;
XX
DT   26-MAR-2002  (first entry)
XX
DE   Drosophila melanogaster polypeptide SEQ ID NO 42489.
XX
KW   Drosophila; developmental biology; cell signalling; insecticide;
KW   pharmaceutical.
XX
OS   Drosophila melanogaster.
XX
PN   WO200171042-A2.
XX
PD   27-SEP-2001.
XX
PF   23-MAR-2001; 2001WO-US009231.
XX
PR   23-MAR-2000; 2000US-0191637P.
PR   11-JUL-2000; 2000US-00614150.
XX
PA   (PEKE ) PE CORP NY.
XX
PI   Venter JC, Adams M, Li PWD, Myers EW;
XX
DR   WPI; 2001-656860/75.
DR   N-PSDB; ABL16002.
XX
PT   New isolated nucleic acid detection reagent for detecting 1000 or more
PT   genes from Drosophila and for elucidating cell signaling and cell-cell
PT   interactions.
XX
PS   Disclosure; SEQ ID NO 42489; 21pp + Sequence Listing; English.
XX
CC   The invention relates to an isolated nucleic acid detection reagent
CC   capable of detecting 1000 or more genes from Drosophila. The invention is
CC   useful in developmental biology and in elucidating cell signalling and
```

```
CC   cell-cell interactions in higher eukaryotes for the development of
CC   insecticides, therapeutics and pharmaceutical drugs. The invention
CC   discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC   sequences (ABL01840-ABL16175) and the encoded proteins (ABB5737-
CC   ABB72072). The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 506 AA;

Query Match            4.6%;  Score 168.5; DB 4;  Length 506;
Best Local Similarity 20.6%;  Pred. No. 8.6e-05;
Matches 114;  Conservative  72;  Mismatches 239;  Indels 129;  Gaps  22;

Qy    84 QHLDSSAANQATEKFVSQMSFMENVRSSAQH--DQRKMVREEEDFAVPVVINSRRSQSHG 141
         |  ::: |  ::  :::  :  :::: ::  :|   |   :
Db     3 QQVENELVNESTQAMTSTTT-QEKVVGSVNNLVEQQCVEMEFETSSDLVVQGSPSEESK 61

Qy   142 RTKSGIEKEKHTPMVAPSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGF 201
         :  |  : :  ::  ||                 :::
Db    62 KVASDASAIVTAPQVQPAPH------VDSSPQASGLSLLAAYSDDSDDEKVTPVQNGD 114

Qy   202 VISLDVSVTE------------EIDLEKSASS------HDRVNDYNASLRQESRNRL 240
         ::|:|                ::||
Db   115 NDVIEVPVTDPASSTTAYRPVVAVSSDDESSKSSSSSDSDSDSEGEYLTVLRKKIDKRI 174

Qy   241 YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQI----- 294
             :            :|::||::
Db   175 ----------NTVDCDEDDEDF-----DEDGATGDRSRRROPPKVRGEMLLDDLPPIHQL 219

Qy   295 ------NEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFW-RARKAIANQRVFAV--EVFDVLG 346
         :
Db   220 EITVPEDECVELGKVQSIVDQLVLVSVLPNSMLFDLDTVLFLEKGRKVLG--EVFDVLG 276

Qy   347 QLFE------------LHRLIKVQKLIAASPDLLLDEISFLGKV---SAKSYPVKKL 388
         |:||                |::
Db   277 QVSDPLYCVRFNSNKQILDRGIKIGDVVYCAPKTEHTQFVILSKLMQVRGSDASW---- 331

Qy   389 LPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVGRLSNQGHHQQS---NYMPFA 444
         |  :                  |  :::|::
Db   332 ---EHDVEPPARYV------DHSDDEEREARAEQRKKRQRDRTNSTDSVDTVTSVATT 381

Qy   445 NNPPASPAPNGYCFPP-------QPPPSGNHQQWLIPVMSPSEGLIYKP-H 487
         |                      |||:|||  ||
Db   382 ATKASSVAP------PPRQRGRRGQRESFRQSQRPSINQHNQ----NQPQDEQYNFHPSYN 432

Qy   488 PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ 547
             |:|| |||  | :|    :  |    | ||  |  : |    : :|
Db   433 PGSWHSNYYQNY--HQAAANFNMAQQHFGMPFFVPVNYGVGMPYAMPPMYPHMYPPPPPFA 490

Qy   548 QQQPNEQMNQFGHP 561
         || |::  |  |
Db   491 PPPPNNQSHQ-GQP 503
```

Search completed: July 29, 2004, 08:22:21
Job time : 103 secs

This Page Blank (uspto)

```
          GenCore version 5.1.6

          Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on:   July 29, 2004, 08:22:26 ; Search time 41 Seconds
                                     (without alignments)
                                     875.123 Million cell updates/sec

Title:           US-10-719-885-2
Perfect score:   3669
Sequence:        1 MKRGKDEEKILEPMFPRLHV.........AARIFQSIQBERKRYDSSKP 695

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters:     389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :       Issued Patents AA:*
                 1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
                 2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
                 3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
                 4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
                 5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
                 6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	* Query Match	Length	DB	ID	Description
1	3669	100.0	695	4	US-09-513-057C-2	Sequence 2, Appli
2	3669	100.0	695	4	US-09-513-057C-35	Sequence 35, Appli
3	3321	90.5	695	4	US-09-234-827B-2	Sequence 4, Appli
4	2806	76.5	595	4	US-09-234-827B-4	Sequence 4, Appli
5	2609.5	71.1	653	4	US-09-513-057C-13	Sequence 13, Appli
6	773.5	21.1	760	4	US-09-513-057C-27	Sequence 27, Appli
7	621.5	16.9	389	4	US-09-513-057C-23	Sequence 23, Appli
8	436.5	11.9	540	4	US-09-513-057C-33	Sequence 33, Appli
9	397	10.8	248	4	US-09-513-057C-29	Sequence 29, Appli
10	281.5	7.7	179	4	US-09-513-057C-24	Sequence 24, Appli
11	231	6.3	49	4	US-09-513-057C-25	Sequence 25, Appli
12	202.5	5.5	185	4	US-09-513-057C-19	Sequence 19, Appli
13	189	5.2	116	4	US-09-513-057C-16	Sequence 16, Appli
14	170.5	4.6	189	4	US-09-513-057C-31	Sequence 31, Appli
15	162	4.4	816	2	US-08-267-803B-9	Sequence 9, Appli
16	162	4.4	816	3	US-09-041-886-17	Sequence 17, Appli
17	158.5	4.3	1061	3	US-08-701-154A-5	Sequence 5, Appli
18	155.5	4.2	1036	3	US-08-891-640-3	Sequence 3, Appli
19	154.5	4.2	1402	4	US-09-125-635-12	Sequence 12, Appli
20	150.5	4.1	928	4	US-09-506-066E-2	Sequence 2, Appli
21	150.5	4.1	2441	1	US-08-194-468-2	Sequence 2, Appli
22	150.5	4.1	2441	3	US-09-961-739-2	Sequence 2, Appli
23	150.5	4.1	2441	4	US-09-514-247A-8	Sequence 8, Appli
24	150.5	4.1	2441	4	US-09-686-316-2	Sequence 2, Appli
25	148.5	4.0	1253	4	US-09-489-039A-12097	Sequence 12097, A
26	148	4.0	1005	2	US-08-935-450-2	Sequence 2, Appli
27	148	4.0	1005	4	US-09-338-123-2	Sequence 2, Appli
28	146.5	4.0	1214	4	US-10-164-595-24	Sequence 24, Appl
29	144	3.9	678	5	PCT-US93-03027-3	Sequence 3, Appli
30	141	3.8	1274	4	US-09-095-443-2	Sequence 2, Appli
31	140.5	3.8	2442	3	US-08-061-376-5	Sequence 10, Appl
32	140.5	3.8	3969	4	US-08-514-247A-10	Sequence 5, Appl
33	140	3.8	720	4	US-09-508-824-11	Sequence 11, Appl
34	140	3.8	1312	4	US-09-041-886-19	Sequence 19, Appl
35	140	3.8	1312	4	US-09-648-281-2	Sequence 2, Appl
36	140	3.8	1312	4	US-09-707-919A-19	Sequence 19, Appl
37	140	3.8	1312	4	US-09-083-268-3	Sequence 3, Appl
38	139	3.8	612	4	US-08-894-454-163	Sequence 163, App
39	137.5	3.7	788	2	US-08-918-914-4	Sequence 4, Appl
40	136.5	3.7	729	4	US-09-625-188-20	Sequence 20, Appl
41	132.5	3.6	1031	4	US-09-914-259-24	Sequence 24, Appl
42	131	3.6	776	4	US-09-266-225D-10	Sequence 10, Appl
43	130.5	3.6	1507	4	US-09-914-259-37	Sequence 37, Appl
44	130.5	3.6	2843	1	US-07-741-940-2	Sequence 2, Appl
45	130.5	3.6	2843	4	US-08-289-548A-2	Sequence 2, Appli
46	130.5	3.6	2843	2	US-08-452-654-2	Sequence 2, Appli
47	130.5	3.6	2843	2	US-08-370-235A-2	Sequence 2, Appli
48	130.5	3.6	703	3	US-08-449-731-2	Sequence 2, Appli
49	128.5	3.5	2842	4	US-08-910-925-4	Sequence 4, Appli
50	128.5	3.5	2842	1	US-07-741-940-7	Sequence 7, Appli
51	128.5	3.5	2842	4	US-08-289-548A-7	Sequence 7, Appli
52	128.5	3.5	2842	2	US-08-452-654-7	Sequence 7, Appli
53	128.5	3.5	2842	2	US-08-449-731-7	Sequence 7, Appli
54	128.5	3.5	2843	4	US-08-452-655B-2	Sequence 2, Appli
55	128.5	3.5	2843	4	US-08-289-548A-7	Sequence 7, Appli
56	128.5	3.5	2843	4	US-08-452-655B-7	Sequence 7, Appli
57	128.5	3.5	2843	3	US-08-450-582-7	Sequence 7, Appli
58	128.5	3.5	2973	3	US-08-821-355A-7	Sequence 7, Appli
59	128.5	3.5	2973	3	US-09-003-687A-7	Sequence 7, Appli
60	128.5	3.5	2973	3	US-09-136-605-7	Sequence 7, Appli
61	128	3.5	949	4	US-10-164-595-65	Sequence 65, Appli
62	128	3.5	1093	3	US-08-545-860D-55	Sequence 55, Appli
63	128	3.5	1093	5	PCT-US94-04496-55	Sequence 55, Appli
64	127	3.5	796	3	US-08-857-076-40	Sequence 40, Appli
65	127	3.5	858	3	US-08-857-076-41	Sequence 41, Appli
66	127	3.5	892	3	US-08-857-076-42	Sequence 42, Appli
67	127	3.5	1404	4	US-08-801-308-1	Sequence 1, Appli
68	127	3.5	1850	4	US-09-620-093A-5	Sequence 5, Appli
69	126.5	3.4	2074	4	US-09-491-356C-9	Sequence 9, Appli
70	126	3.4	748	3	US-09-725-459B-27	Sequence 27, Appli
71	126	3.4	2023	4	US-09-491-356C-8	Sequence 8, Appli
72	125	3.4	1395	3	US-09-540-245A-15	Sequence 15, Appli
73	124.5	3.4	380	2	US-09-026-587-4	Sequence 4, Appli
74	124.5	3.4	380	2	US-09-227-420-4	Sequence 4, Appli
75	124.5	3.4	380	4	US-09-387-811-4	Sequence 4, App
76	124.5	3.4	635	4	US-09-417-197-125	Sequence 125, App
77	124.5	3.4	753	3	US-08-725-159B-25	Sequence 25, Appl
78	124.5	3.4	1184	4	US-09-266-225D-18	Sequence 18, Appl
79	124.5	3.4	1185	3	US-09-041-886-23	Sequence 23, Appl
80	124.5	3.4	2414	1	US-08-227-536-2	Sequence 2, Appl
81	124.5	3.4	2382	5	PCT-US95-04682-2	Sequence 2, Appli
82	124	3.4	460	4	US-09-056-556-184	Sequence 184, App
83	124	3.4	460	4	US-09-072-596-179	Sequence 179, App
84	124	3.4	460	4	US-09-072-967-184	Sequence 184, App
85	124	3.4	1096	3	US-08-415-946-3	Sequence 3, Appli
86	124	3.4	1464	3	US-08-891-640-2	Sequence 2, Appli
87	124	3.4	2185	4	US-09-854-856-36	Sequence 36, Appli
88	124	3.4	2245	4	US-09-854-856-4	Sequence 4, Appli
89	124	3.4	2322	4	US-09-854-856-34	Sequence 34, Appli
90	124	3.4	2382	4	US-09-854-856-2	Sequence 2, Appli
91	123.5	3.4	716	4	US-09-219-983A-20	Sequence 20, Appl
92	123.5	3.4	1155	1	US-08-094-948A-29	Sequence 29, Appl
93	123.5	3.4	1155	5	PCT-US96-09319-29	Sequence 29, Appl
94	123	3.4	2157	4	US-09-854-856-52	Sequence 52, Appli
95	123	3.4	2217	4	US-09-854-856-20	Sequence 20, Appli
96	123	3.4	2294	4	US-09-854-856-50	Sequence 50, Appli
97	123	3.4	2354	4	US-09-854-856-18	Sequence 18, Appli
98	122.5	3.3	747	3	US-08-973-273-3	Sequence 3, Appli
99	122.5	3.3	2004	4	US-09-854-856-58	Sequence 58, Appli
100	122.5	3.3	2064	4	US-09-854-856-26	Sequence 26, Appli

ALIGNMENTS

RESULT 1
US-09-513-057C-2
; Sequence 2, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-513-057C-2

  Query Match   100.0%; Score 3669; DB 4; Length 695;
  Best Local Similarity   100.0%; Pred. No. 7.4e-308;
  Matches 695; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60

Qy   61 NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120

Qy   121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180

Qy   181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240

Qy   241 YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300

Qy   301 DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360

Qy   361 IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420

Qy   421 SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480

Qy   481 GLIYKPHPGMAHTGHYGGYYGHYGMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPY 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GLIYKPHPGMAHTGHYGGYYGHYGMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPY 540

Qy   541 CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600

Qy   601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTVQTTRDGGGVTR 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTVQTTRDGGGVTR 660

Qy   661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
           |||||||||||||||||||||||||||||||||||
Db   661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695

RESULT 2
US-09-513-057C-35
; Sequence 35, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-513-057C-35

  Query Match   100.0%; Score 3669; DB 4; Length 695;
  Best Local Similarity   100.0%; Pred. No. 7.4e-308;
  Matches 695; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60

Qy   61 NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61 NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120

Qy   121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180

Qy   181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240

Qy   241 YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241 YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300

Qy   301 DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301 DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360

Qy   361 IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361 IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420

Qy   421 SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421 SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480

Qy   481 GLIYKPHPGMAHTGHYGGYYGHYGMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPY 540
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481 GLIYKPHPGMAHTGHYGGYYGHYGMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPY 540

Qy   541 CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   541 CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600

Qy   601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTVQTTRDGGGVTR 660
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTVQTTRDGGGVTR 660

Qy   661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
           |||||||||||||||||||||||||||||||||||
Db   661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695

```
Db    644  TRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP  695
```

RESULT 3
US-09-234-827B-2
; Sequence 2, Application US/09234827B
; Patent No. 6448471
; GENERAL INFORMATION:
; APPLICANT: Puzio, Piotr S.
; APPLICANT: Grundler, Florian M.W.
; TITLE OF INVENTION: Nematode feeding structure specific gene and its
; TITLE OF INVENTION: application to produce nematode resistant plants
; FILE REFERENCE: U-012084-2
; CURRENT APPLICATION NUMBER: US/09/234,827B
; CURRENT FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: US 60/072,142
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 695
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-234-827B-2

Query Match   90.5%; Score 3321; DB 4; Length 695;
Best Local Similarity 91.0%; Pred. No. 7.7e-278;
Matches   648; Conservative   4; Mismatches 26; Indels 34; Gaps   3;

```
Qy    1  MKRGKDEEKILEPMFPRLHVNDADKGPGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN  60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||| |||||
Db    1  MKRGKDEEKILEPMFPRLHVNDADKGPGPRAPPRNKMALYEQLSIPSQRFGDHGTRNSRSN  60

Qy   61  NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61  NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV  120

Qy  121  REEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  REEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL  180

Qy  181  ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL  240
         |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db  181  ATCSKPEVRDQVKANRRSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL  240

Qy  241  YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD  300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD  300

Qy  301  DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL  360
         ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db  301  DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIAKQQRVFAVQLFELHRLIKVQKL  360

Qy  361  IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  361  IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES  420

Qy  421  SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQ-----------  469
         ||||||||||||||||||
Db  421  SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQ            469

Qy  470  --QWLIPVMSPSEGLI----YKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPG  523
              :: :  :                ||||||||||||||||||||||||||||||||||
Db  464  LLQEIISNGDPCNVSLGRTDIQPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPG  523

Qy  524  NGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQ  583
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  524  NGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQ  583

Qy  584  QQPTKSYPRARKSRQGSTGSSPSGPGQISGSKSFGPFAAVDEDSNINNAPEQTMTTTTT  643
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  584  QQPTKSYPRARKSRQGSTGSSPSGPGQISGSKSFGPFAAVDEDSNINNAPEQTMTTTTT  643

Qy  644  TRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP  695
         |||||||||||||||||||||||||||||||||||||||||||||||||||
```

RESULT 4
US-09-234-827B-4
; Sequence 4, Application US/09234827B
; Patent No. 6448471
; GENERAL INFORMATION:
; APPLICANT: Puzio, Piotr S.
; APPLICANT: Grundler, Florian M.W.
; TITLE OF INVENTION: Nematode feeding structure specific gene and its
; TITLE OF INVENTION: application to produce nematode resistant plants
; FILE REFERENCE: U-012084-2
; CURRENT APPLICATION NUMBER: US/09/234,827B
; CURRENT FILING DATE: 1999-01-21
; PRIOR APPLICATION NUMBER: US 60/072,142
; PRIOR FILING DATE: 1998-01-22
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 4
; LENGTH: 595
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-234-827B-4

Query Match   76.5%; Score 2806; DB 4; Length 595;
Best Local Similarity 89.9%; Pred. No. 1.6e-233;
Matches   550; Conservative   4; Mismatches 24; Indels 24; Gaps   3;

```
Qy  101  QMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSS  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    1  QMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSS   60

Qy  161  HHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSAS  220
         ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||
Db   61  HHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANRRSGGFVISLDVSVTEEIDLEKSAS  120

Qy  221  SHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDR  280
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  121  SHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDR  180

Qy  281  EYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQ  340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||| |
Db  181  EYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQ  240

Qy  341  QRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLP  400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  241  QRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLP  300

Qy  401  HVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPP  460
         ||||||||||||||||||||||||||||||||||||||
Db  301  HVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPP  360

Qy  461  QPPPSGNHQ------------QWLIPVMSPSEGLI----YKPHPGMAHTGHYGGYYGHY  503
         :: :  :
Db  344  QQTTHRLRLQMDIAFLLSLLQEIISNGDPCNVSLGRTDIQPHPGMAHTGHYGGYYGHY  403

Qy  504  MPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGN  563
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  404  MPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGN  463

Qy  564  LQNTQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPGQISGSKSFGPFAAV  623
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  464  LQNTQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPGQISGSKSFGPFAAV  523

Qy  624  DEDSNINNAPEQTMTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSI  683
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  524  DEDSNINNAPEQTMTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSI  583

Qy  684  QEERKRYDSSKP  695
         ||||||||||||
Db  584  QEERKRYDSSKP  595
```

```
RESULT 5
US-09-513-057C-13
; Sequence 13, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
;   LENGTH: 653
;   TYPE: PRT
;   ORGANISM: Cardamine oligosperma
US-09-513-057C-13

  Query Match             71.1%;  Score 2609.5;  DB 4;  Length 653;
  Best Local Similarity   76.7%;  Pred. No. 1.7e-216;
  Matches 517;  Conservative 47;  Mismatches 47;  Indels 43;  Gaps 12;

Qy     1  MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHG--TMNSR  58
          ||||||:|||||||||||||||||||||||||||||||||||||||||||||||  :||
Db     1  MKRGKDDEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGNLSLSSR  60

Qy    59  SNNTSTLVHPGPSS-QPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQR  117
          |||||||||||||  : ||||||||||||||||| :  :||||: ||||||:::|:|||
Db    61  SNNTSTLVHPGPSNQQSCGVERNLSVQHLDSSAAVHVTENFVSQMPFMENVMRSLAKHDQR  120

Qy   118  KMVREEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQN  177
          |:||||:||||| :||||:|||| |||| :|||||
Db   121  KTVREEDDFAVPVFVNSRRFQSHGSTKSGIVIEKHT  156

Qy   178  VCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESR  237
          :||||||  |:||:||:|||||||   |::||||:||||||||| |||||:|::|||||||
Db   157  -TLATCSK-LVRDKVKMNAKSGGFI---DLSSTEEVDLEKSASSYDRVNDCNSSLRQESR  211

Qy   238  NRLYRDGGKTRLKDTDNGAEBSHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEE  297
          |:|||||: : ||| ||  ||||| |:||:|||||||||:||: |||::||  |||||||
Db   212  NKLYRDGGEAMKDTANRVESHLVTESHSEEGHGSPDDDNGHEYCRSRGGVSLQQINEE  271

Qy   298  ASDDVSDDSMVDSISSIDVSPPDVVGILGQKRFWRARKAIANQRVFAVQLFELHRLIKV  357
          ||||||| ||||||||||||||||| ||||||||||||| |||:|||||| ||||||||
Db   272  ASDDVSDNSMVDSISSIDVSPPDVVGALGQKRFWRARKAITNQQRVFAVQLPELHRLIKV  331

Qy   358  QKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVV--KQRGDSEKTDQ  415
          |:|||||||::||||:::|||||||||||||:|||| ||||||:||  ||: ||||||
Db   332  QRLIAASPDIVLDEINVIGKVSAKSYPVKKLVPSEFTVKPPLPQVVVNKQHRSDSEKTDQ  391

Qy   416  HKMESSAENVVGRLSNOGH---HQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWL  472
          ||||:|||||||||||| |   |||||:||| |:||||||:||| :||||||||||-|||||
Db   392  HKMECSAENVVGRLSNOQHHHNHQPSNVMPFPSNPPASPAVNGCCYPPQPPPSGN-QQWL  450

Qy   473  IPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHMPGMGFPPPYPYGM  532
          ||||||||||||||||||||||||||||||:||| |||||: ||||||||||    |||||
Db   451  IPVMSPSEGLIYKPHPGMGHTGHYGGYYGHFMPPPMVMPPFHPGMGFPPPPGNGYFPPYGV  510

Qy   533  MPTIMNPYCSSQQQQQ-QPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYP  591
          :|::||||  |||||| | |:||:|||| |||| : |:: :|:||||| :::||  |||||
Db   511  IPAMMNPYGPGQQQQQQQPQANEQTNQFGYSGNLQNNTHQES-SVNEAAPPQEPL-TKSYP  568

Qy   592  RARKSRQSTGSPSGPQGISGSKSFRPFAAVDEDSNINN--APEQTMTTTTTTTTTTRTTVT  649
          ||||||| | |:|||||||||:|||||:|||| : ||:  || | |||||||||||||||||
Db   569  RARKSRQVSTASSASGREGISGSTSFRPFSAVDEDNDNNNDAPDQMTTTTTTTTTTTTRTTVT  628

Qy   650  QTTRDGGVTRVIK  663
          |||||| ::|||||
Db   629  QTTRDGGEVTRVIK  642

RESULT 6
```

```
US-09-513-057C-27
; Sequence 27, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
;   LENGTH: 760
;   TYPE: PRT
;   ORGANISM: Oryza sativa
US-09-513-057C-27

  Query Match             21.1%;  Score 773.5;  DB 4;  Length 760;
  Best Local Similarity   32.1%;  Pred. No. 4.9e-58;
  Matches 261;  Conservative 105;  Mismatches 257;  Indels 191;  Gaps 36;

Qy     4  GKDEE-KILEPMFPRLHVNDADK-GGPRAPPRNKMALYEQLSIPSQRF----GDHGTMNS  57
          || :: : :|||||||||||: :| ||||||||||||||||| ::|::||    ::|: :
Db    12  GKEAKGKVMGPLFPRLHVNDAAKGGGPRAPPRNKMALYEQFTVPSHRFSGGGGGVGGS  71

Qy    58  RSNNTSTLVHPGPSSQPCGVERNL------------SVQHLDSSAANQATEKFVSQMSF  104
          ::|:||
Db    72  PAHSTSAASQSQSQSQVYGRDSSLFQPFNVPSNRPGHSTEKINSDKINKKISGSRKELGM  131

Qy   105  MEN------VRSSAQHDQR-------KMVREEDFAVPVYINSRRSQSHGRTK  144
          ||:        :|: :
Db   132  LSSQTKGMDIYASRSTAEAPQRRAENTIKSSSGKRLADDDEFMVPSVFNSRFPQYSTQEN  191

Qy   145  SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQV-CLATCSKPEVRDQVK-ANARSGGFV  202
          :| : ||:||:|| :::         |: :|: |    : :|:||: : | |
Db   192  AGVQ-DQSTPLVAANPHKS-------PSTVSKSSTKCYNTVSKKLERIHVSDVKSRTPLKD  244

Qy   203  ISLDVSVT-EEIDLEKSASSHDRVNDYNASLRQESRN-RLYRDGGKTR-LKDTD---NGA  256
          :::   : :|:  |:||:| ||| :         : | :         ||
Db   245  KEMEAAQTSKNVEVEKSSSFHASKDMF----------ESRHAKVYPKMDKTGIINDSEPHGGN  298

Qy   257  ESHLATENHS--------------QEGHGSPEDIDNDREY------SKSRACASLQQIN  295
          . :: :|:            :  : | |
Db   299  SGHQATSRNGGSMKFQNPPMRRNEISSNPSSENTDRHYNLPQGGIEETGTKRKRLLEQHD  358

Qy   296  EEASDDVS----------DDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQ  341
          .||||||          ||||||||||||||:|||||| |||||||||||| ||||
Db   359  AEKSDDVSRLLEQHDAENIDDVSDSSVECITGWEISPDKIVGAIGTKHFWKARRAIMNQQ  418

Qy   342  RVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPH  401
          ||||||:|||| |:||||||||||| :|  ::  |                :|:
Db   419  RVFAVQVFELHKLVKVQKLIAASPHVLIESDPCLGNALLGS---KNKLVEENLKAQPLLV  475

Qy   402  VVVKQRGDSEKTQHKMESSAENV---------VGRLSNQGHHQQSNYMPFAN---NPPA  449
                  :      :: :|:|          : : : :            :: |::
Db   476  ATI---DVEPSLQQPEVSKENTEDSPPSPHDTGLGSGQRDQAATNGVSKSNRRATPVA  531

Qy   450  SP-APNGYCFPQPQPPSGNHQQWLIPVMSPSEGLIYKPHPG----------------M  490
          |  : |:  : ||  |||:|||||:|||||||:|||:||| |                |
Db   532  SDNKQNNWGVQLQPP----QNQWLVPVMSPLEGLVYKPYSGPCPPAGSIIAPFYANCTPL  587

Qy   491  AHTGHYGGYYGHYMPTPMVMPQYHPGMGFPP--PGNGYFPYGMMPTIMNPYCSSQQQQQ  548
          : : :::    |::| :|| || :     : :                  ||:|
Db   588  SLPSTAGDFMNSAYGVPMHPQPQHMGAPGPPSMEMN-YFPPFSI-PVMNP  635

Qy   549  QQPNEQMNQFGHPGNLQ---NTQQQQRSDNEPAPQQQQOPTKSYP----RARKSR  597
               ::: : :| : :    ::         :     :           |||||
Db   636  TAPAPVVEQGRHPSMPQPYGNFEQQSWISCN------MSHPSGIWRFHASRDSEA  684

Qy   598  QGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTTTRTTVTQTTRDGGG  657
          | :: :|   : : :   |||  : :    ::      :                 : |:
Db   685  QASSASSPFDRFQCGGS---GPVSAFPTVSAQNNQPQP-----------SYSSRD-N  726
```

RESULT 7
US-09-513-057C-23
; Sequence 23, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-513-057C-23

Query Match   16.9%;   Score 621.5;   DB 4;   Length 389;
Best Local Similarity   39.0%;   Pred. No. 2.4e-45;
Matches 164; Conservative   58; Mismatches 132; Indels   67; Gaps   17;

Qy 297 EASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIK 356
              |||||||||| | | | : ||||||||||| ||: | | | : ||||||||
Db   2 DRGDDLSETSRVESISGTDISPDDIVGIIGLKRFWKARRAIVNQQRVFAIQVFELHRLIK   61

Qy 357 VQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQH 416
              ||||||||||| | |||: || |||||||||||| |||||||||| | | |: ||
Db   62 VQRLIAGSPNSSLEDPAYIGK-PLKSSSIKR-LPLDCIVRE--SQSVLKRKHDSEK-PHF 116

Qy 417 KMESSAARNVVGRLS---NQGHHQQSNVMPFANNPPASPAPNG----YCFPPQPPPSGNH 468
              :||:|:|:|| || || :|||| |||||||||||| ||||| |||| |
Db 117 RMEHTAESNVGKASLSTVQNGSQLSSHKPFSGTPLPTPVTNDSNAGPWCF---QQPSGH- 172

Qy 469 QQWLIPVMSPSEGLIYKPHPGMAHTG----------HYGGYYGHYMPTPMVMPQYHPG   516
              |||||||||||||||||||||| | | |
Db 173 -QWLIPVMSPSEGLVYKPFSGPGPTSPICGSGPGSGSSPTMGNFFAPTYGVPAPNPHYQ-G 230

Qy 517 MGF---PPPGNGYFPPYGMGMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQN-------T 567
              || |||||||||: || | |:||: || : | ||
Db 231 MGVPFAPPTGHGYFRQYGM--PAMNPYISS------TASEESNQYTMPG-LQHQFSGVVD 281

Qy 568 QQQQRDNEPAPQQQQQPTKSYPRARKSRQGST-GSSPSGPQGISGSKSFRPFAAVDED 626
              ||||||||||||||||||||||||||| | :| | :| ||| | || |
Db 282 DVQHSTSGLSNVLNQKKRNVPDVVRYQSTKDNEVQASSASSP-IETAGRNMLSLFPTSPVT 341

Qy 627 SNINNAPQTMTTTTTTRTTVTVQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEE   686
              :|| | || | | | | | | | ||||||||||||||||||||||
Db 342 DNRDGSPQACVPDNP-----------------ARVIKVVPHNARSATESVARIFQSIQQE   384

Qy 687 R   687
          |
Db 385 R   385

RESULT 8
US-09-513-057C-33
; Sequence 33, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33

; LENGTH: 540
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-513-057C-33

Query Match   11.9%;   Score 436.5;   DB 4;   Length 540;
Best Local Similarity   23.6%;   Pred. No. 3.5e-29;
Matches 177; Conservative   98; Mismatches 195; Indels 279; Gaps   32;

Qy   5 KDEEK--ILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLS------IPSQRFGDHGTMN   56
          ||| | : | || | |||: | || | | | | : || | |:
Db   5 KDEAKRITIPLFPRVHVNDTGRGG----------------------LSQQFDGKTMSLVSSKRPNLPSPTN   54

Qy   57 SRSNNTSTL--VHPGPSSQPC-GVERNLSVQHLDSSAANQATEKFVS----QMSFMEN 107
              : |: | | || ||| | :|::| || | | | |:| || |||
Db   55 NISDSLSTFSLSLPPPNNARLIDGPEKNQFSPIYNTKFEGKLNKKGINYTSPKGSSVTN 114

Qy 108 VRSSA--QHDQRKMVREEDFAVPYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIR 165
          | |:| ||:| |: | | | | || | | |:: :
Db 115 TKPSSIKQNEYLKNLTSLDSIKSPIVIHSE----------IDPQANTDL-----SLQ 156

Qy 166 FQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRV 225
          | : | |:   : ||
Db 157 F------------------CTSGSSKP----------GEAVVGSKILLSE--------RL 181

Qy 226 NDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKS 285
                                | ||||:|||| || : | ||| || || |
Db 182 EDEN-------------------QNGSPNVMKTQSYRR----NFAEFNNETQ-KKP 213

Qy 286 RACASLQQINEEASDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFA 345
          || :| | |:|||||||||||||||||||||||||||| || | | | |||||
Db 214 KTLPRREQV------------ASNCSAIELSLGISASSYDIARVIGEKRFWKMRTYMINQQKIFA 266

Qy 346 VQFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVK 405
          ||:||||| || || | | | |||| | || |: | |||| ||| | ||||||||| :
Db 267 GQVFELHRLIMVQKMVAKSPNLFLES--------------------------------K 293

Qy 406 QRGDSEKT-DQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGY------- 456
           ::|:|: |:| | |:||||| :| || ||||||||||| || |||
Db 294 LNGVKHGTMRSSHQLAMAASKV--RKPNTENHK-----------PVPEEYPEHMKPK 337

Qy 457 ---------------CFPPQ-PPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGH 502
                          ||| ||||| |||||||||||||||||||||| | |||||
Db 338 LPLPSISKELVTPIWPQQLLPPPGN--QWLVPVITDSDGLVYKPPPG- 382

Qy 503 YMPTP----MVMPQY----------HPGMGFP--PPGNGYFPPYGMMPTIMNPYCSSQQQQQ 549
             | :| || ||| | |||||||| | | |: :|||||||
Db 383 -PCPPSSAAFMVPVYGQDSLETPFRFPVSSPFSHSYFPPNARTTV----------- 427

Qy 550 QPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQP--TKSYPRARKSRQGSTGSSPS 606
          | || ||| | |::   ::| | |:| | :| | | | || || || ::|
Db 428 ---DQTNPFG----------QFQRWSNTSSHMTQAIPFSLKKSQESNDSDIHGSTASSP- 473

Qy 607 GPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTT----RDGGGVTRVI 662
          | ::| | | | |||| | :| | ||| | | | | | |||| ||
Db 474 ----------------------------PERHKLEVLPLFPTEPTHQTDEYKQKQQPMLRAI 507

Qy 663 KVVPHNAKLASENAARIFQSIQEERKRYD 691
          |:||||| | || |||||| || ||||||
Db 508 KAVPHNSTSASESAARIFRFIQEERRDSD 536

RESULT 9
US-09-513-057C-29
; Sequence 29, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/513,057C
; CURRENT FILING DATE: 2000-02-24
; NUMBER OF SEQ ID NOS: 35
; SEQ ID NO 35

Qy 658 VTRVIKVVPHNAKLASENAARIFQSIQEERKRYD 691
              |||||||||||:| || |||||||| || ||||
Db 727 QTNVIKVVPHNSRTASESAARIFRSIQMERQRDD 760

```
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 29
;   LENGTH: 248
;   TYPE: PRT
;   ORGANISM: Zea mays
;   FEATURE:
;   NAME/KEY: UNSURE
;   LOCATION: (88)..(88)
;   OTHER INFORMATION:
;   NAME/KEY: unsure
;   LOCATION: (98)..(98)
;   OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
;   NAME/KEY: unsure
;   LOCATION: (113)
;   OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
;   NAME/KEY: unsure
;   LOCATION: (113)..(113)
;   OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
;   NAME/KEY: unsure
;   LOCATION: (114)..(114)
;   OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
US-09-513-057C-29

Query Match             10.8%;  Score 397;  DB 4;  Length 248;
Best Local Similarity   29.5%;  Pred. No. 2.9e-26;
Matches   121;  Conservative  34;  Mismatches  75;  Indels 180;  Gaps  13;

Qy    297 EASDDVSDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIK 356
              :|::| :|::: ::||||||: :: :||||||||||||||||||||||||
Db      3 EQNDLSDSS-VESLPGMEISPDDVVSAIGPKHFWKARRAIVNQQRVFAVQVFELHRLIK  61

Qy    357 VQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQH 416
              |||||||||||||| :|::  |||  :  |||
Db     62 VQKLIAASPHVLIEGDPCLGKSLAVSXK----------------------------  89

Qy    417 KMESSAENVVGRLSNQGHHQQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVM 476
                                                              QWLIPXM  99
Db     90 ------------RLS-----------------------------------------

Qy    477 SPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHP--------GMGFPPPGNG---- 525
              :|||| :  ::|||:  |:  |:|
Db    100 SPFEGLVYKPYPGXXPSG------GSLLAPPFFAS--YPTSSSSTAGGDFMSSACGARLM 151

Qy    526 ----YFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPCNLQNTQQQQQRSDNEPAPQ 581
              ||||| ||| : :           :  |:
Db    152 SAPVFPSFS-MPAV---------SGSAVEQVSHVA---ASQHKRNSCSEAVLA 192

Qy    582 QQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQIMTTTT 641
              :
Db    193 SRDSEV--------QGSSASSPASSE---------------------------------- 210

Qy    642 TTTRTTVTQTTRDGGGVTRVIKVPHNAKLASENARAIFQSIQEERKRYD 691
              | : ||:|||||||:   ::||
Db    211 -----TAAQ------PRVIRVPHTARTASESAARIFRSIQMERKQND 247
```

```
RESULT 10
US-09-513-057C-24
;   Sequence 24, Application US/09513057C
;   Patent No. 6433251
;   GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 24
;   LENGTH: 179
;   TYPE: PRT
;   ORGANISM: Lycopersicon esculentum
US-09-513-057C-24

Query Match             7.7%;  Score 281.5;  DB 4;  Length 179;
Best Local Similarity   41.1%;  Pred. No. 1.7e-16;
```

```
Matches    79;  Conservative  25;  Mismatches  61;  Indels  27;  Gaps   9;

Qy      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN  60
              ||| |::|  : | :|::  || :|   :||||||||:|||||||||:| :   |::
Db      1 MKRGTGEEKVMGPMFPRLNVNDTEKGGPRAPPRNKMALYEQLSIPSQRY-NPGDLPHNSS  59

Qy     61 NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSF----MENVRSSAQHDQ 116
              :: ::|:||: :   ::|  :  :|  :  |:|:    |:|
Db     60 NSANIVLPHP-SQENEHERGVLFSR-QLPALRHPVEKPYGRSSGSNTPLREVKSKRQ--- 114

Qy    117 RKVVREEEDFAVPVVINSR----RSQSHGRTKSGIEKE---KHTPMVAPSSHHSIRFQE 168
              : :|| ||  :||  ::      :  :|||| ||:||     |::| : :  :|:::
Db    115 ---TEKEDFRVPTFDNSKERAVNTEDYSKGTSDIDKRDSTLKRTDQL--SHVTPRENL 167

Qy    169 VNQTGS--KQNV 178
              |: ::  |::|
Db    168 VNTFGESHKTNI 179
```

```
RESULT 11
US-09-513-057C-25
;   Sequence 25, Application US/09513057C
;   Patent No. 6433251
;   GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 25
;   LENGTH: 49
;   TYPE: PRT
;   ORGANISM: Lycopersicon esculentum
US-09-513-057C-25

Query Match             6.3%;  Score 231;  DB 4;  Length 49;
Best Local Similarity  85.7%;  Pred. No. 5.8e-13;
Matches    42;  Conservative   4;  Mismatches   3;  Indels   0;  Gaps   0;

Qy      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRF 49
              ||| |::|| : |:||||||||||  ||||||||:|||||||||||||
Db      1 MKRGKGEEKLMGPMFPRLHVNDTEKGGPKAPPRNKMALYEQLSIPSQRF 49
```

```
RESULT 12
US-09-513-057C-19
;   Sequence 19, Application US/09513057C
;   Patent No. 6433251
;   GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
;   SEQ ID NO 19
;   LENGTH: 185
;   TYPE: PRT
;   ORGANISM: Glycine max
US-09-513-057C-19

Query Match             5.5%;  Score 202.5;  DB 4;  Length 185;
Best Local Similarity  33.3%;  Pred. No. 1.1e-09;
Matches    69;  Conservative  20;  Mismatches  69;  Indels  49;  Gaps   9;

Qy    503 YMPTPMVMPQYHPGMG----FPPPGNGYFPPYGVMGPMPTIMNPYCSSQQQQQQPNEQMNQF 558
              :  :  :  :|||     | :::    |::::   |   :|:|           ||:
Db     10 FMNPAYQFPPASHPVVGVSPFVPPASHTYFAPFGM--PVMN------QATSGSAVEQVNQF  61

Qy    559 GHPGNL-QN----------TQQQQQRSDNEPAPQQQQQPTKSYPRARKSR--QGSTGSSP 605
```

Db   62 AAQGSHGQNGHSSVEGADFNTHHNQSSSNLPVQKNGARLHVKKSQALKERGLQGSTRSSP 121

Qy   606 SG-PQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRDGGGVTRVIKV 664

Db   122 SEMAQGIRARKI----------ADGSDAQSLHADETRQQ-----------TQAIKV 158

Qy   665 VPHNAKLASENAARIFQSIQEERKRYD 691

Db   159 VPHNRKSATFSAARIVQSIQEERKQHD 185

RESULT 13
US-09-513-057C-16
; Sequence 16, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 16
;   LENGTH: 116
;   TYPE: PRT
;   ORGANISM: Pisum sativum
US-09-513-057C-16

Query Match   5.2%;   Score 189;   DB 4;   Length 116;
Best Local Similarity 46.7%;   Pred. No. 8.4e-09;
Matches   56;   Conservative   14;   Mismatches   34;   Indels   16;   Gaps   7;

Qy   353 RLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEK 412
          ||||||:||||| |||| :|:|| :::|||: :|||| | : ||| : :||||:|||
Db   1 RLIKVQQLIAGSPDLLFDDGAFLGK-SLPDGSTPKKLSLEYVVKARLQN-LKRKVDSEK 57

Qy   413 TDQHKMESSAENVVGRLS-----NQGHHQQSNYMPFANNPPAS--PAPNG---YCFPPQP 462
          |||| :|: |||: |::   : | :|| : | : | |||:|
Db   58 INQN-MECSAENAVGKTSISSVKNTSH-LSSSMPFAGNPHQGNVAADNGMGPWCFNQSP 114

RESULT 14
US-09-513-057C-31
; Sequence 31, Application US/09513057C
; Patent No. 6433251
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/513,057C
;   CURRENT FILING DATE: 2000-02-24
;   NUMBER OF SEQ ID NOS: 35
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 31
;   LENGTH: 189
;   TYPE: PRT
;   ORGANISM: Zea mays
US-09-513-057C-31

Query Match   4.6%;   Score 170.5;   DB 4;   Length 189;
Best Local Similarity 29.2%;   Pred. No. 6.7e-07;
Matches   63;   Conservative   23;   Mismatches   77;   Indels   53;   Gaps   11;

Qy   501 GHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQ-----QQPNEQ 554
          ||:| |::| :|| || ::| | :   :| || |:   : :|
Db   3 GHMVPP---------GAPAMHMNYFPPFSM--PVMNPGTPASAVEQSHAAAPQPHGH 49

Qy   555 MNQ------FGHPGNLQNTQQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS------- 600
          |::| ::|:: ||
Db   50 MDQQSLISCNMSHPSGVWRFLASR---DSEP--QASSATSPFDRLQVQGDGSAPLSFFP 103

Qy   601 TGSSPS-GPQGISGSKSFRPFAAVDEDSN---INNAPEQTMTTTTTTTRTTVTQTTRDGG 656
          |||:|| |: || |||| | :: |:   :|||| :: ||:::||||:||
Db   104 TASAPNVQPPSSGGR--------DRDQQNHVIRVVPRNAQTASVPKAQPQPSSGGRD-- 153

Qy   657 GVTRVIKVVPHNAKLASENAARIFQSIQEERKRYDS 692
          ||:|||||||||||||||:|||||||||||||:||
Db   154 QKNHVIRVVPHNAQTASESAAWIFRSIQMERNQNDS 189

RESULT 15
US-08-267-803B-9
; Sequence 9, Application US/08267803B
; Patent No. 5834183
; GENERAL INFORMATION:
;   APPLICANT: Orr, Harry T.
;   APPLICANT: Ranum, Laura P.W.
;   APPLICANT: Chung, Ming-yi
;   APPLICANT: Zoghbi, Huda Y.
;   TITLE OF INVENTION: Gene Sequence for Spinocerebellar Ataxia
;   Type 1 and Method for Diagnosis
;   NUMBER OF SEQUENCES: 85
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Mueting, Raasch, Gebhardt & Schwappach, P.A.
;   STREET: P.O. Box 581415
;   CITY: Minneapolis
;   STATE: MN
;   COUNTRY: USA
;   ZIP: 55458-1415
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/267,803B
;   FILING DATE: 28-JUN-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: McCormack, Myra H.
;   REGISTRATION NUMBER: 36,602
;   REFERENCE/DOCKET NUMBER: 110.00030120
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 612-305-1217
;   TELEFAX: 612-305-1228
; INFORMATION FOR SEQ ID NO: 9:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 816 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-267-803B-9

Query Match   4.4%;   Score 162;   DB 2;   Length 816;
Best Local Similarity 23.0%;   Pred. No. 3e-05;
Matches   85;   Conservative   32;   Mismatches   130;   Indels   122;   Gaps   17;

Qy   397 PPLPHVVVKQRGDSEK-----TDQHKMESSA--ENVVGRLSNQGHHQQSNYMPFANNPP 448
          | | ||:|:|| |: ||:|||::| || | |:::|| |||| ||
Db   14 PKKREIPATSRSSEEKAPTLPSDNHRVEGTAWLPGNPGGRGHGGGRH---------GPA 63

Qy   449 ASPAPNGV--------CFPPQPPPSGNHQQWLIPVMPSEGLLYKPHPG---- 489
          ||: | :   : |||| |:|| ||: | | :
Db   64 GTSVELGLQQGIGLHKALSTGLDYSPPSAPRS------VPVATLPAAYATPQPGTPVS 116

Qy   490 --MAHTGH-------YGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPP---- 529
                          | ::| |: |:| : ||| ||| ||||
Db   117 PVQYAHLPHTFQFIGSSQYSGTYASFIPSQLIPPTANPVTSAVASAAGATTPSQRSQLEA 176

Qy   530 YGMMPTIMNP------YCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQR------S 574
          || :|::||| ||
Db   177 YSTLLANMGSLSQTPGHKAEQQQQQQQQQQQQQHQQQQQQQQQQQHLSRAPGLIT 236

Qy   575 DNEPAPQQQQQPTKSYPRARKSRQ--GSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNA 632
                :: : :::   :::   .:   ::
Db   237 PGSPPPAQQNQ----YVHISSSPQNTGRTASPPAIP------------------------VHLH 272

Qy   633 PEQTMT--TTTTTRRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQ------ 684
                :   ::::::::: : : ::   ::: ::: :: :::::: :::
Db   273 PHQTMIPHLTLGPPSQVVMQYADSGS-----HFVPREATKKAE-SSRLQQAIQAKEVLN 326

Qy   685 ---EERKRY 690
              :: :||
Db   327 GEMEKSRRY 335

RESULT 17
US-08-701-154A-5
; Sequence 5, Application US/08701154A
; Patent No. 6380373
; GENERAL INFORMATION:
; APPLICANT: O'Malley, Bert W.
; APPLICANT: Tsai, Ming-Jer
; APPLICANT: Tsai, Sophia Y.
; APPLICANT: Onate, Sergio A.
; TITLE OF INVENTION: STEROID RECEPTOR COACTIVATOR
; TITLE OF INVENTION: COMPOSITIONS AND METHODS OF USE
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Lyon & Lyon
; STREET: 633 West Fifth Street
; STREET: Suite 4700
; CITY: Los Angeles
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 90071-2066
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 3.5" Diskette, 1.44 Mb
; MEDIUM TYPE: storage
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: IBM P.C. DOS 5.0
; SOFTWARE: FastSEQ for Windows 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/701,154A
; FILING DATE: August 21, 1996
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/003,784
; FILING DATE: September 15, 1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Warburg, Richard J.
; REGISTRATION NUMBER: 32,327
; REFERENCE/DOCKET NUMBER: 220/243
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (213) 489-1600
; TELEFAX: (213) 955-0440
; TELEX: 67-3510
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1061 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-701-154A-5

Query Match   4.3%; Score 158.5; DB 4; Length 1061;
Best Local Similarity   20.1%; Pred. No. 8.9e-05;
Matches 156; Conservative 89; Mismatches 240; Indels 291; Gaps 37;

Qy   18 LHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDH------------GTMNSRSNNTS 63
               : ::   :   ::   : : :::   :   : ::: : :::
Db   233 LHNNDRLSDGDSKYSQTSHKLVQLITTAEQLRHADIDTSCKDVLSCTGTSNSASANSS 292

Qy   64 TLVHPG---PSSQPCGVERNLSVQHLDSAANQATEKFVSQMSFMENVRSSAQHDRQKMV 120
               ::: || :   :::   :   :::   : :
Db   [continued]

RESULT 16
US-09-041-886-17
; Sequence 17, Application US/09041886
; Patent No. 6235872
; GENERAL INFORMATION:
; APPLICANT: Bredesen, Dale E.
; APPLICANT: Rabizadeh, Sharroz
; TITLE OF INVENTION: Proapoptotic Peptides, Dependence
; TITLE OF INVENTION: Polypeptides and Methods of Use
; NUMBER OF SEQUENCES: 72
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Campbell & Flores LLP
; STREET: 4370 La Jolla Village Drive, Suite 700
; CITY: San Diego
; STATE: California
; COUNTRY: United States
; ZIP: 92122
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/041,886
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Campbell, Cathryn A.
; REGISTRATION NUMBER: 31,815
; REFERENCE/DOCKET NUMBER: P-LJ 2626
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (619) 535-9001
; TELEFAX: (619) 535-8949
; INFORMATION FOR SEQ ID NO: 17:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-041-886-17

Query Match   4.4%; Score 162; DB 3; Length 816;
Best Local Similarity   23.0%; Pred. No. 3e-05;
Matches 85; Conservative 32; Mismatches 130; Indels 122; Gaps 17;

Qy   397 PPLPHVVKQRGDSEK-----TDQHKMESSA----ENVVGRLSNQGHHQQSNYMPFANNPP 448
               : :: ::: :   : : ::: : :   : :   ::
Db   14 PKKREIPATSRSSEEKAPTLPSDNHRVEGTAWLPGNPGGRGHGGGRH---------GPA 63

Qy   449 ASPAPNGY-------------CFPPQQPPSGNHQQWLIPVMSPSEGLIYKPHPG----- 489
               : :::   :   : :   :::   : :::   :::::
Db   64 GTSVELGLQQGIGLHKALSTGLDYSPPSAPRS------VPVATTLPAAYATPQPGTPVS 116

Qy   490 ---MAHTGH------------YGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPP----- 529
                  :: :   : ::: ::   :   :   : : : :
Db   117 PVQYAHLPHTFQFIGSSQYSGTYASFIPSQLIPPTANPVTSAVASAAGATTPSQRSQLEA 176

Qy   530 YGMMPTIMNP------YCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQQR--------S 574
               : : :   : :||||||||||||:|:|||||||||||||||||||::
Db   177 YSTLLANMGSLSQTPGHKAEQQQQQQQQQQQQQHQQQQQQQQQQQHLSRAPGLIT 236

CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/021,247
FILING DATE: 12-JUL-1996
ATTORNEY/AGENT INFORMATION:
NAME: Steffe, Eric K.
REGISTRATION NUMBER: 36,688
REFERENCE/DOCKET NUMBER: 1383.0130001/EKS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1036 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: not relevant
MOLECULE TYPE: peptide

US-08-891-640-3

Query Match   4.2%;   Score 155.5; DB 3;   Length 1036;
Best Local Similarity 19.9%; Pred. No. 0.00016;
Matches 156; Conservative 89; Mismatches 240; Indels 297; Gaps 37;

```
Qy    18 LHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDH----------------GTMNSRSNNTS    63
            :                              :  ::    :   ::
Db   233 LHNNDRLSDGDSKYSQTSHKLVQLLITTABQLRHADIDTSCKDVLSCTGTSNSASANSS   292

Qy    64 TLVHPG--PSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV   120
            ::        :::       ::        :   :     : :  :
Db   293 -----GGSCPSS-------------HSSLTARHKILHRLLQEGSPSDITTLSVEPDKKDSA   335

Qy   121 REEEDFAVPVYINS--------------RRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEV   169
            ||                        ||
Db   336 STSVSVTGVQGNSSIKLELDASKKKESKDHQLLRYLLDKDEKDLRSTP----NLSLDDV   391

Qy   170 NQTGSKQ-------NVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEID-----LE   216
            ||                                             ::  :  ::
Db   392 KVKVEKKEQMDPCNTNPTPMTKATPE--EIKLEAQS-QFTADLD----QFDQLLPTLE   442

Qy   217 KSAS----SHDRVND-----------------YNASLRQESR-NRLYRDGGKTR   248
            :::                           ||  ||
Db   443 KAAQLPGLCETDRMDGAVTSVTIKSEILPASLQSATARPTSRLNRL----PELE   498

Qy   249 LKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMV   308
            ||
Db   499 LEAIDN-------QFGQPGTGDQIPWTNN----------------------------

Qy   309 DSI---SSIDVSPDDVVGILGQKRFWRARKAIANQQRVF-----AVQLFELHRLIKVQK   359
            ::                          ||              ||
Db   540 DELLCPPTTVEGRNDE---------------------------KALLEQLVSFLSGKDETELAELDRALGIDK   585

Qy   360 LI-----------AASPDLLLDE------------------ISFLGKVSAK---   381
            ||                ||                         ||
Db   586 LVQGGGLDVLSERFPPQQATPPLIMEERPNLYSQPYSSPFPTANLPSPFQGMVRQKPSLG   645

Qy   382 SYPVK------------KLLPSEFLVKPPLP----HVVVKQR--GDSEKTDQHKM----E   419
            :||                 ||                ||
Db   646 TMPVQVTPPRGAFSPGMGMQPRQTLNRPPAAPNQLRLQLQQRLQQQQLIHQNRQAILNQ   705

Qy   420 SSAENVVGRLSNQGHHQQSNYMP-------------------------------   442
            ||
Db   706 FAATAPVGINMRSGMQQQITPQPPLNAQMLAQRQRELYSQQHRQRLIQQQRAMLMRQQS   765

Qy   443 FANN-PPAS----------PAPNGYCFPPQ-------PPPSGNHQQWLIPVMSPSEG   481
            ||                                 ||
Db   766 FGNNLPPSSGLPVQTGNPRLPQGAPQQFYPPNYGTNPGTPPAS---TSPFSQ   815

Qy   482 LIYKPHPGMAH--------TGHYGGYYGHYMPTPMVMPQY-HPGMGFPPPGNGYFPP-   529
            :|
Db   816 LAANPEASLANRNSMVSRGMTGNIGGQFGTGINPQMQQNVFQYPGAGMVPQGEANFAPSL   875

Qy   530 --YGMMPTIMNPYCSSQQQQ-----QQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAP   580
```

RESULT 18
US-08-891-640-3
; Sequence 3, Application US/08891640
; Patent No. 6268173
; GENERAL INFORMATION:
APPLICANT: Chambon, Pierre
APPLICANT: Gronemeyer, Hinrich
APPLICANT: Voegel, Johannes
APPLICANT: Lutz, Yves
TITLE OF INVENTION: Transcriptional Intermediary Factor-2
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sterne, Kessler, Goldstein & Fox P.L.L.C.
STREET: 1100 New York Avenue, NW, Suite 600
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/891,640
FILING DATE: Herewith

```
Db     876  SPGSSMVEMPIPPQSSLLQQTPPASGYQSPDMKAWQQGAIGNNNVFSQAVQ---NQPTP  932
            :|: ||     ||   ||  |||         |: |: |||
Db    1330  --------------------------QPTPMYQPSDMKGWPSGNLARNGSPQQQFAPQG  1363

Qy     581  QQ  582
                QQ
Db     933  AQ  934


RESULT 19
US-09-125-635-12
; Sequence 12, Application US/09125635
; Patent No. 6562589
; GENERAL INFORMATION:
;   APPLICANT: THE UNITED STATES OF AMERICA represented by THE SE
;   TITLE OF INVENTION: AIB1, A novel steriod receptor co-activator
;   FILE REFERENCE: 49944
;   CURRENT APPLICATION NUMBER: US/09/125,635
;   CURRENT FILING DATE: 1998-08-21
;   PRIOR APPLICATION NUMBER: 60/049,728
;   PRIOR FILING DATE: 1997-06-17
;   NUMBER OF SEQ ID NOS: 12
;   SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
;   LENGTH: 1402
;   TYPE: PRT
;   ORGANISM: Mus musculus
US-09-125-635-12

Query Match        4.2%;   Score 154.5;  DB 4;  Length 1402;
Best Local Similarity   20.6%;  Pred. No. 0.00029;
Matches    134;  Conservative   72;  Mismatches  226;  Indels  217;  Gaps   31;

Qy      16  PRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSNNTSTLVHPGSSQPC   75
            ||:|     ::  || :  | :|| :       :  ||       :
Db     878  PRMMDSQENYGANMGPNRNVPV-----NPTSSPGDWGLANSRASRMEPL----ASSPL  926

Qy      76  GVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEDFAVPVYINSR  135
            |       :  :||      :|     ::|  |: |||| :  |||||||||||||||
Db     927  GRTGADYSATLPRPAMGGSVPTLPLRSNRLPGARPSLQQQQQQQQQQQQQQQQQQQQQQ  986

Qy     136  RSQSH------GRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLA------T  182
            :               ||: :: |: :   |   |  ||  :
Db     987  QQQQQQMLQMRTGEIPMGMGVNPYSPAV-PS-----------NQPGSWPEGMLSMEQGPHGS  1035

Qy     183  CSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYR  242
            ::|||  |    |       ||::   :  :||||||||||||||||||
Db    1036  QNRPLLRN----------SLD-------DLLGPPSNAEGQSDERALLDQ--LHTLLS  1073

Qy     243  DGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQI--NEEASD  300
            ||| :|||  :   :  |   ||        |:  |  :  :  :
Db    1074  NTDATGLEEIDRAL------------GIPELV--------NQGQALESKQDVFQGQEAA  1112

Qy     301  DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL  360
                         |||||||||||||    |  || :|  |::|||  ::  | |  :||
Db    1113  ------------------VMMDQKAALYGQ----------------TYPAQGPPLQGFNIQ-  1140

Qy     361  IAASPDLLLDEISFLGKVSAK-SYPVKKLLPSEFLVKPPLP-------HVVKQR------  407
            :||||||||||||:|||     ||||||||||||||     | ||:|
Db    1141  -GQSPSF----NSMMGQISQQGSFPLQGMHPRAGLVRPRTNTPKQLRMQLQGQQRLQGQQFL  1195

Qy     408  GDSEKTDQHKMESSAENVVGR--------LSNQ-----GHHQQSNYMPFANNPPAS  450
            |::|||||||||||||||||          :|       ||
Db    1196  NQSRQALEMKMENPAGTA'YMRPMMPQAFFNAQMAAQQKRELMSHHLQQQRMAMM---MSQ  1252

Qy     451  PAPNGYCFPPQPPPSGNHQQWLIPVMSPS-EGLIYKPHPGMAHTGHYGYGYYGHYMPTPMV  509
            |||  |         ||   |||       :|:||||
Db    1253  PQPQAF-----SPPPN-------VTASPSMDGVL---AGSA----------------MPQA  1282

Qy     510  MPQYHP------GMGFPPP---GNGYFPPYGMGMPTIMNPYCSSQQQQQQQPNEQMNQFGHP  561
            ||:|:|      |:| |||   ||   |||||||  |||||||||||||  :|:|   ||
Db    1283  PPQQFPYPANYGMGQPPEPAFGRGSSPSAMMSSRMGP--------SQNAMVQHP--------  1329

Qy     562  GNLQNTQQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPQG  610
```

Qy   637 MTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEE 686
         :: :   : :   :.::   : . : :::
Db   851 STPSKVVSTDTLINLNDDVG- -EAVAEAVTEGGKQSLE-AEESGQQVEVE 897

RESULT 21
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
;   APPLICANT: Montminy, Marc R.
;   TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
;   TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
;   TITLE OF INVENTION: RESPONSIVE GENES
;   NUMBER OF SEQUENCES: 3
;   CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
;   STREET: 444 South Flower Street, Suite 2000
;   CITY: Los Angeles
;   STATE: California
;   COUNTRY: USA
;   ZIP: 90071
;   COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/194,468
;   FILING DATE: 10-FEB-1994
;   CLASSIFICATION: 435
;   ATTORNEY/AGENT INFORMATION:
;   NAME: Reiter, Stephen E.
;   REGISTRATION NUMBER: 31,192
;   REFERENCE/DOCKET NUMBER: P41 9672
;   TELECOMMUNICATION INFORMATION:
;   TELEPHONE: (619)-546-4737
;   TELEFAX: (619)-546-9392
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;   LENGTH: 2441 amino acids
;   TYPE: amino acid
;   TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-194-468-2

  Query Match   4.1%; Score 150.5; DB 1; Length 2441;
  Best Local Similarity   24.3%; Pred. No. 0.0014;
  Matches   75; Conservative   24; Mismatches 115; Indels 115; Gaps 11;

Qy   353 RLIKVQKLIAASPDLLLDEISFLGKVSA- -KSYPVKKLLPSEFLVKPPLPHVVKQRGDS 410
         : .:| : : :: :| || .|| | :| :|| .| : : | :
Db 2053 RMPNVQPNRSISPSALQDLLRTLKSPSSPQQQQQVLNILKS- - - -NPQLMAAFIKQR- - - 2105

Qy   411 EKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQ 470
                                    |: | :
Db 2106 - TAKYV- ANQPGMQPQPGLQSQPGMQPQPGMEQQ 2137

Qy   471 WLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPY 530
                ::||| .| | :
Db 2138 - - - - - -PSLQNLNAMQAG- - - - - - - - - - - - - - -VPRPGVPP- - - - - - - - - -PQPAMGGLNPQ 2168

Qy   531 GMMPTIMNP- -YCSSQQQQQQPNEQMNQFGHPGNLQ-N 566
         | : ||||| : ||||| |
Db 2169 GQALNIMNPGHNPNMTNMNPQYREMVRRQLLQHQQQQQQQQQQQQQQNSASLAGGMAGH 2228

Qy   567 TQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDED 626
         : | .| |||||||| | | | :| | | | ||
Db 2229 SQFQQPQGPGGYAPAMQQQRMQQHLPIQGSSMGQM- - -AAPMGQLGQMG- - - -QPGLGADST 2283

Qy   627 SNINNAPEQ 635
         |:|: | :
Db 2284 PNIQQALQQ 2292

RESULT 22
US-08-961-739-2
; Sequence 2, Application US/08961739A
; Patent No. 6061583
; GENERAL INFORMATION:
;   APPLICANT: Montminy, Marc R.
;   TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
;   FILE REFERENCE: SALK1650-1
;   CURRENT APPLICATION NUMBER: US/08/961,739A
;   CURRENT FILING DATE: 1997-10-31
;   EARLIER APPLICATION NUMBER: US 194,468
;   EARLIER FILING DATE: 1994-02-10
;   NUMBER OF SEQ ID NOS: 4
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
;   LENGTH: 2441
;   TYPE: PRT
;   ORGANISM: Mus
;   FEATURE:
;   NAME/KEY: VARIANT
;   LOCATION: (1)...(2441)
;   OTHER INFORMATION: Xaa = Any Amino Acid
US-08-961-739-2

  Query Match   4.1%; Score 150.5; DB 3; Length 2441;
  Best Local Similarity   24.3%; Pred. No. 0.0014;
  Matches   75; Conservative   24; Mismatches 115; Indels 115; Gaps 11;

Qy   353 RLIKVQKLIAASPDLLLDEISFLGKVSA- -KSYPVKKLLPSEFLVKPPLPHVVKQRGDS 410
         : .:| : : :: :| || .|| | :| :|| .| : : | :
Db 2053 RMPNVQPNRSISPSALQDLLRTLKSPSSPQQQQQVLNILKS- - - -NPQLMAAFIKQR- - - 2105

Qy   411 EKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQ 470
                                    |: | :
Db 2106 - TAKYV- ANQPGMQPQPGLQSQPGMQPQPGMEQQ 2137

Qy   471 WLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPY 530
                ::||| .| | :
Db 2138 - - - - - -PSLQNLNAMQAG- - - - - - - - - - - - - - -VPRPGVPP- - - - - - - - - -PQPAMGGLNPQ 2168

Qy   531 GMMPTIMNP- -YCSSQQQQQQPNEQMNQFGHPGNLQ-N 566
         | : ||||| : ||||| |
Db 2169 GQALNIMNPGHNPNMTNMNPQYREMVRRQLLQHQQQQQQQQQQQQQQNSASLAGGMAGH 2228

Qy   567 TQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDED 626
         : | .| |||||||| | | | :| | | | ||
Db 2229 SQFQQPQGPGGYAPAMQQQRMQQHLPIQGSSMGQM- - -AAPMGQLGQMG- - - -QPGLGADST 2283

Qy   627 SNINNAPEQ 635
         |:|: | :
Db 2284 PNIQQALQQ 2292

RESULT 23
US-09-514-247A-8
; Sequence 8, Application US/09514247A
; Patent No. 6365361
; GENERAL INFORMATION:
;   APPLICANT: TANABE SEIYAKU CO. LTD.
;   APPLICANT: TANIGUCHI, Tomoyasu
;   APPLICANT: MIZUKAMI, Junko
;   TITLE OF INVENTION: METHOD FOR IDENTIFYING OR SCREENING AGONIST AND ANTAGONIST TO PPI
;   FILE REFERENCE: TANIGUCHI=6
;   CURRENT APPLICATION NUMBER: US/09/514,247A
;   CURRENT FILING DATE: 2000-02-28
;   PRIOR APPLICATION NUMBER: PCT/JP98/03734
;   PRIOR FILING DATE: 1998-08-24
;   PRIOR APPLICATION NUMBER: JP231084/1997
;   PRIOR FILING DATE: 1997-08-27
;   NUMBER OF SEQ ID NOS: 10

```
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 8
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: mouse
US-09-514-247A-8

Query Match          4.1%;  Score 150.5; DB 4;  Length 2441;
Best Local Similarity 24.3%; Pred. No. 0.0014;
Matches     75;  Conservative  24; Mismatches 115;  Indels  95;  Gaps  11;

Qy    353  RLIKVQKLIAASPDLLLDEISFLGKVSA--KSYPVKKLLPSEFLVKPPLPHVVVKQRGDS  410
               |:|::||   :|||||:      :|  ||       :|   ||
Db   2053  RMPNVQPNRSISPSALQDLLRTLKSPSSPQQQQVLNILKS----NPQLMAAFIKQR---  2105

Qy    411  EKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPQPPPSGNHQQ  470
                                              :    ::
Db   2106  -----------------------------------TAKYV--ANQPGMQPQPGLQSQPGMQPQPGMHQQ  2137

Qy    471  WLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTMVMPQYHPGMGFPPPGNGYFPPY  530
                                                                  :|:|
Db   2138  --------------------------------PSLQNLNAMQAG----------------------VPRPGVPP-----  2168

Qy    531  GMMPTIMNP----------------------------------YCSSQQQQQPNEQMNQFGHPGNLQ-N  566
               ||||
Db   2169  GQALNIMNPGHPNPNMTNMNPQYREMVRRQLLQHQQQQQQQQQQQQQNSASLAGGMAGH  2228

Qy    567  TQQQQQRSDNEPAPQQQQPTKSYPRARKSRQSTGSSPSGPQGISGSKSFRPFAAVDED  626
               |:||| |||  ||    ||  |  :::|
Db   2229  SQFQQPQGPGGYAPAMQQQRMQQHLPIQGSSMGQM-AAPMGQLGQMG-----QPGLGADST  2283

Qy    627  SNINNAPEQ  635
               :|:|
Db   2284  PNIQQALQQ  2292
```

```
RESULT 24
US-09-686-316-2
; Sequence 2, Application US/09686316
; Patent No. 6646115
; GENERAL INFORMATION:
;  APPLICANT: Montminy, Marc R.
;  TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
;  FILE REFERENCE: SALK1650-1
;  CURRENT APPLICATION NUMBER: US/09/686,316
;  CURRENT FILING DATE: 2000-10-10
;  PRIOR APPLICATION NUMBER: US/08/961,739
;  PRIOR FILING DATE: 1997-10-31
;  PRIOR APPLICATION NUMBER: US 194,468
;  PRIOR FILING DATE: 1994-02-10
;  NUMBER OF SEQ ID NOS: 4
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: Mus
; FEATURE:
;  NAME/KEY: VARIANT
;  LOCATION: (1)...(2441)
;  OTHER INFORMATION: Xaa = Any Amino Acid
US-09-686-316-2

Query Match          4.1%;  Score 150.5; DB 4;  Length 2441;
Best Local Similarity 24.3%; Pred. No. 0.0014;
Matches     75;  Conservative  24; Mismatches 115;  Indels  95;  Gaps  11;

Qy    353  RLIKVQKLIAASPDLLLDEISFLGKVSA--KSYPVKKLLPSEFLVKPPLPHVVVKQRGDS  410
               |:|::||   :|||||:      :|  ||       :|   ||
Db   2053  RMPNVQPNRSISPSALQDLLRTLKSPSSPQQQQVLNILKS----NPQLMAAFIKQR---  2105

Qy    411  EKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPQPPPSGNHQQ  470
                                              :    ::
Db   2106  -----------------------------------TAKYV--ANQPGMQPQPGLQSQPGMQPQPGMHQQ  2137
```

```
; SEQ ID NO 8
; LENGTH: 2441
; TYPE: PRT
; ORGANISM: mouse
US-09-514-247A-8
(continued)

Qy    471  WLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTMVMPQYHPGMGFPPPGNGYFPPY  530
                                                                  :|:|
Db   2138  --------------------------------PSLQNLNAMQAG----------------------VPRPGVPP-----  2168

Qy    531  GMMPTIMNP----------------------------------YCSSQQQQQPNEQMNQFGHPGNLQ-N  566
               ||||
Db   2169  GQALNIMNPGHPNPNMTNMNPQYREMVRRQLLQHQQQQQQQQQQQQQNSASLAGGMAGH  2228

Qy    567  TQQQQQRSDNEPAPQQQQPTKSYPRARKSRQSTGSSPSGPQGISGSKSFRPFAAVDED  626
               |:||| |||  ||    ||  |  :::|
Db   2229  SQFQQPQGPGGYAPAMQQQRMQQHLPIQGSSMGQM-AAPMGQLGQMG-----QPGLGADST  2283

Qy    627  SNINNAPEQ  635
               :|:|
Db   2284  PNIQQALQQ  2292
```

```
RESULT 25
US-09-489-039A-12097
; Sequence 12097, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
;  APPLICANT: Gary Breton et. al
;  TITLE OF INVENTION:  NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA
;  TITLE OF INVENTION:  PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE:  2709.2004001
;  CURRENT APPLICATION NUMBER: US/09/489,039A
;  CURRENT FILING DATE:  2000-01-27
;  PRIOR APPLICATION NUMBER:  US 60/117,747
;  PRIOR FILING DATE:  1999-01-29
;  NUMBER OF SEQ ID NOS:  14342
; SEQ ID NO 12097
; LENGTH: 1253
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-12097

Query Match          4.0%;  Score 148.5; DB 4;  Length 1253;
Best Local Similarity 22.7%; Pred. No. 0.00082;
Matches    102;  Conservative  37; Mismatches 162;  Indels 149;  Gaps  25;

Qy    334  RKAIANQQRVFA------------------------VQLFELHRLIKVQKLI---AASPDLLLDEISFLGKVS  379
               :|:||:                        :|:
Db     92  RSALARRKLAEKFTNPMGRKTDAALFSGKRMDDGEEVVQYSASGAPVAADDVLFSGASA  151

Qy    380  AKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQ---------------HKM-------------ESS  421
               :|                    ||:
Db    152  ARPAEDDVLFSGASAVRP------------GDFDPYDPLLNGHSIAEPVSAAAAATAAPQAW  201

Qy    422  AENVVGRLSNQGHH------------------QQSNVMPFANNPPASPAPNGY  456
               |:|                         :|
Db    202  AESPV------GHHGAAPAYQPEASYPPGQAYQPEPAPFQQAAYQP---PAGQTAPQAY  251

Qy    457  CFPQPPPSGNHQQWLIPVMSPSEG----LIYKPHPGMAHTGHYGGYYGH---YM-P  505
               ||:                    :|:
Db    252  --QPEPAP---YQQ--PVVDPRVGQPAPQAYQPEPAPYQQPAPYYAGQPAPQAYQPEP  303

Qy    506  TPMVMPQYHPGMGFPPPGNGYFP--PYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPG  562
               |:                    |:|
Db    304  APYQQPAYDPYAGQPAP-QAYQPEPAPY-QQPT-YDPHAGQPAPQAYQP-------EPA  352

Qy    563  NLQNT--------QQQQQRSDNEPAPQQQ------QQPTKSYPRARKSRQSTGSS  604
               :||                           |:
Db    353  PYQQPAYDPYAGQPAPQTYQPEPAYQQPAYDPHAGQPAPQPYQPEPAAYQPQSAPVPPP  412

Qy    605  PSGPQGISGSKSFRP---FAAVDEDSNINN------APEQTMTTTTTTRRTTVTQ  650
               |                            :|
Db    413  EPEPEVVQEEVKRPPLYFEEVEEKRARERELLASWYQPIPEPESPIATKLTPPTTASK  472

Qy    651  TTRDGGGVTRVIKVVP---HNAKLASENAA  677
               |:                     :|:
Db    473  PPVE---TTVISAVAAGVEQATAASGGAA  498
```

This Page Blank (uspto)

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 08:23:21 ; Search time 50 Seconds
                                   (without alignments)
                                   4360.193 Million cell updates/sec

Title: US-10-719-885-2
Perfect score: 3669
Sequence: 1 MKRGKDEEKILEPMFPRLHV.........AARIFQSIQEERKRYDSSKP 695

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched: 1291235 seqs, 313682936 residues

Total number of hits satisfying chosen parameters: 1291235

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 100 summaries

Database :   Published Applications AA:*
    1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
    2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
    3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
    4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
    5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
    6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
    7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
    8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
    9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
   10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
   11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
   12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
   13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
   14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
   15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
   16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
   17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
   18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3669	100.0	695	9	US-09-746-801A-2	Sequence 2, Appli
2	3669	100.0	695	9	US-09-746-801A-35	Sequence 35, Appl
3	2609.5	71.1	653	9	US-09-746-801A-13	Sequence 13, Appl
4	904	24.6	324	9	US-09-746-801A-61	Sequence 61, Appl
5	777	21.2	599	9	US-09-746-801A-59	Sequence 59, Appl
6	773.5	21.1	760	16	US-09-746-801A-27	Sequence 27, Appl
7	773.5	21.1	760	16	US-10-437-963-136891	Sequence 136891,
8	725.5	19.8	534	12	US-10-424-599-230456	Sequence 230456,
9	686.5	18.7	781	16	US-10-437-963-131862	Sequence 131862,
10	621.5	16.9	389	9	US-09-746-801A-23	Sequence 23, Appl
11	495.5	13.5	168	9	US-09-746-801A-63	Sequence 63, Appl
12	459.5	12.5	317	9	US-09-746-801A-57	Sequence 57, Appl
13	436.5	11.9	540	9	US-09-746-801A-33	Sequence 33, Appl
14	397	10.8	248	9	US-09-746-801A-29	Sequence 29, Appl
15	393.5	10.7	335	12	US-10-424-599-222485	Sequence 222485,
16	377	10.3	334	12	US-10-425-114-42323	Sequence 42323, A
17	316	8.6	383	12	US-10-425-114-63567	Sequence 63567, A
18	306	8.3	213	12	US-10-424-599-253615	Sequence 253615,
19	295	8.0	203	12	US-10-424-599-280208	Sequence 280208,
20	281.5	7.7	179	9	US-09-746-801A-24	Sequence 24, Appl
21	256.5	7.0	110	12	US-10-424-599-285624	Sequence 285624,
22	255.5	7.0	89	12	US-10-424-599-247729	Sequence 247729,
23	231	6.3	49	9	US-09-746-801A-25	Sequence 25, Appl
24	212	5.8	208	12	US-10-425-114-43416	Sequence 43416, A
25	202.5	5.5	185	9	US-09-746-801A-19	Sequence 19, Appl
26	195	5.3	795	12	US-10-424-599-174901	Sequence 174901,
27	193.5	5.3	802	15	US-10-369-493-22462	Sequence 22462, A
28	189.5	5.2	1130	14	US-10-032-585-7758	Sequence 7758, Ap
29	189	5.2	116	9	US-09-746-801A-16	Sequence 16, Appl
30	188	5.1	115	9	US-09-746-801A-47	Sequence 47, Appl
31	187	5.1	522	14	US-10-425-114-62797	Sequence 62797, A
32	186.5	5.1	1138	14	US-10-074-475-194	Sequence 194, App
33	181	4.9	631	12	US-10-424-599-266585	Sequence 266585,
34	180	4.9	42	9	US-09-746-801A-37	Sequence 37, Appl
35	176.5	4.8	1132	12	US-10-425-114-62671	Sequence 62671, A
36	176.5	4.8	1291	12	US-10-312-352-32	Sequence 32, Appl
37	173	4.7	618	12	US-10-424-599-174900	Sequence 174900,
38	172.5	4.7	2321	12	US-10-193-874-17	Sequence 17, Appl
39	172.5	4.7	2358	12	US-10-193-874-14	Sequence 14, Appl
40	172.5	4.7	2540	12	US-10-193-874-16	Sequence 16, Appl
41	172.5	4.7	2552	12	US-10-193-874-15	Sequence 15, Appl
42	170.5	4.6	189	9	US-09-746-801A-31	Sequence 31, Appl
43	167.5	4.6	1424	16	US-10-437-963-160683	Sequence 160683,
44	166.5	4.5	1243	15	US-10-408-765A-1964	Sequence 1964, Ap
45	166.5	4.5	1441	15	US-10-355-218-2	Sequence 2, Appli
46	165	4.5	780	9	US-09-770-689A-5	Sequence 5, Appli
47	163	4.4	1965	15	US-10-369-493-3279	Sequence 3279, Ap
48	162	4.4	407	12	US-10-425-114-59366	Sequence 59366, A
49	162	4.4	407	16	US-10-437-963-147959	Sequence 147959,
50	162	4.4	717	12	US-10-425-114-58581	Sequence 58581, A
51	161	4.4	816	14	US-10-207-706-3	Sequence 3, Appli
52	161	4.4	971	14	US-10-093-524-8	Sequence 8, Appli
53	160.5	4.4	2017	14	US-10-114-270-86	Sequence 86, Appli
54	160	4.4	1441	16	US-10-437-963-192225	Sequence 192225,
55	159	4.3	699	16	US-10-437-963-191304	Sequence 191304,
56	158.5	4.3	1272	14	US-10-032-585-7777	Sequence 7777, Ap
57	158.5	4.3	459	16	US-10-381-247B-2	Sequence 2, Appli
58	157.5	4.3	1494	14	US-10-029-386-32987	Sequence 32987, A
59	157	4.3	326	16	US-10-282-122A-49391	Sequence 49391, A
60	156	4.3	1586	12	US-09-925-300-1444	Sequence 1444, Ap
61	156	4.3	531	12	US-10-114-605-93	Sequence 93, Appli
62	156	4.3	1403	12	US-10-381-247B-17	Sequence 17, Appl
63	155.5	4.2	1450	9	US-09-842-256-3	Sequence 3, Appli
64	155.5	4.2	1036	16	US-09-320-797-355	Sequence 355, Ap
65	154.5	4.2	1367	16	US-10-408-765A-919	Sequence 919, App
66	153.5	4.2	1644	14	US-10-379-616-12	Sequence 12, Appl
67	153	4.2	1402	14	US-10-032-585-7784	Sequence 7784, Ap
68	153	4.2	1381	9	US-09-850-887-4	Sequence 4, Appli
69	152.5	4.2	274	16	US-10-177-293-106	Sequence 106, App
70	151.5	4.1	883	15	US-10-104-047-2505	Sequence 2505, Ap
71	151.5	4.1	1101	15	US-10-094-749-2342	Sequence 2342, Ap
72	151	4.1	756	15	US-10-425-114-60327	Sequence 60327, A
73	150.5	4.1	799	15	US-10-408-765A-1141	Sequence 1141, Ap
74	150.5	4.1	626	16	US-10-112-372-4	Sequence 4, Appli
75	150.5	4.1	492	16	US-10-408-765A-2180	Sequence 2180, Ap
76	150.5	4.1	827	16	US-10-109-886-8	Sequence 8, Appli
77	150	4.1	1313	13	US-10-116-275-179	Sequence 179, Appl
78	150	4.1	2441	15	US-10-389-566-2372	Sequence 2372, Ap
79	149.5	4.1	1572	15	US-10-425-114-62454	Sequence 62454, A
80	149.5	4.1	1791	12	US-10-192-985-1	Sequence 1, Appli
81	149	4.1	620	14	US-10-425-114-43565	Sequence 43565, A
82	148.5	4.0	643	14	US-10-112-372-6	Sequence 6, Appli
83	148.5	4.0	479	16	US-10-437-963-106955	Sequence 106955,
84	148	4.0	813	13	US-10-262-445-66	Sequence 66, Appl
85	148	4.0	1566	15	US-10-114-091-2	Sequence 2, Appli
86	148	4.0	1001	15	US-10-262-445-62	Sequence 62, Appl
87	148	4.0	1005			
88	148	4.0	1005			

```
89    148    4.0    1068   15   US-10-262-445-64         Sequence 64, Appl
90    148    4.0    1446   12   US-10-343-649-7          Sequence 7, Appli
91    147.5  4.0    425    12   US-10-425-114-57692      Sequence 57692, A
92    147.5  4.0    427    12   US-10-425-114-45447      Sequence 45447, A
93    147.5  4.0    759    12   US-10-282-1222A-71823    Sequence 71823, A
94    147.5  4.0    1360   16   US-10-473-574-12         Sequence 12, Appl
95    147    4.0    589    12   US-10-424-599-251769     Sequence 251769,
96    147    4.0    976    15   US-10-144-194A-92        Sequence 92, Appli
97    146.5  4.0    638    14   US-10-038-010-4          Sequence 4, Appli
98    146.5  4.0    793    14   US-10-112-372-8          Sequence 8, Appli
99    146.5  4.0    1206   15   US-10-085-198-144        Sequence 144, App
100   146    4.0    1267   12   US-10-092-900A-128       Sequence 128, App
```

ALIGNMENTS

RESULT 1
US-09-746-801A-2
; Sequence 2, Application US/09746801A1
; Patent No. US20020083494A1
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/746,801A
;   CURRENT FILING DATE: 2000-12-20
;   NUMBER OF SEQ ID NOS: 68
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
;   LENGTH: 695
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-746-801A-2

Query Match   100.0%; Score 3669; DB 9; Length 695;
Best Local Similarity   100.0%; Pred. No. 1.7e-271;
Matches   695; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1   MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60

Qy   61   NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120

Qy   121   REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121   REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180

Qy   181   ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181   ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240

Qy   241   YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241   YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300

Qy   301   DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301   DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360

Qy   361   IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361   IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420

Qy   421   SAENVVGRLSNQGHHQQSNYMPFANNPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421   SAENVVGRLSNQGHHQQSNYMPFANNPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480

Qy   481   GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPYGMGFPPPYGMGFPPPYGMGMPTIMNPY 540
```

RESULT 2
US-09-746-801A-35
; Sequence 35, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
;   APPLICANT: Wagner, et al.
;   TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
;   FILE REFERENCE: 1505-54357
;   CURRENT APPLICATION NUMBER: US/09/746,801A
;   CURRENT FILING DATE: 2000-12-20
;   NUMBER OF SEQ ID NOS: 68
;   SOFTWARE: PatentIn version 3.1
; SEQ ID NO 35
;   LENGTH: 695
;   TYPE: PRT
;   ORGANISM: Arabidopsis thaliana
US-09-746-801A-35

Query Match   100.0%; Score 3669; DB 9; Length 695;
Best Local Similarity   100.0%; Pred. No. 1.7e-271;
Matches   695; Conservative   0; Mismatches   0; Indels   0; Gaps   0;

Qy   1   MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1   MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60

Qy   61   NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   61   NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120

Qy   121   REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   121   REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180

Qy   181   ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   181   ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRL 240

Qy   241   YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   241   YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD 300

Qy   301   DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   301   DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL 360

Qy   361   IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   361   IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES 420

Qy   421   SAENVVGRLSNQGHHQQSNYMPFANNPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   421   SAENVVGRLSNQGHHQQSNYMPFANNPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE 480

Qy   481   GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPYGMGFPPPYGMGMPTIMNPY 540
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   481   GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPYGMGFPPPYGMGMPTIMNPY 540
```

```
Db    481   GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPYGMGFPPPYGMGMPTIMNPY  540

Qy    541   CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS  600
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541   CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS  600

Qy    601   TGSSPSGPQGISGSKSFRFPAAVDEDSNINNAPEQTMTTTTTTRTTVTQTTRDGGVTR  660
                |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601   TGSSPSGPQGISGSKSFRFPAAVDEDSNINNAPEQTMTTTTTTRTTVTQTTRDGGVTR  660

Qy    661   VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP  695
                ||||||||||||||||||||||||||||||||||
Db    661   VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP  695
```

```
Qy  541 CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600
        |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  541 CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS 600

Qy  601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRDGGGVTR 660
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  601 TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRDGGGVTR 660

Qy  661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
        |||||||||||||||||||||||||||||||||||
Db  661 VIKVVPHNAKLASENAARIFQSIQEERKRYDSSKP 695
```

RESULT 3
US-09-746-801A-13
; Sequence 13, Application US/09746801A
; Patent No. US2002008349A1
; GENERAL INFORMATION:
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 13
; LENGTH: 653
; TYPE: PRT
; ORGANISM: Cardamine oligosperma
US-09-746-801A-13

Query Match   71.1%; Score 2609.5; DB 9; Length 653;
Best Local Similarity 76.7%; Pred. No. 1.5e-190;
Matches 517; Conservative 47; Mismatches 47; Indels 43; Gaps 12;

```
Qy   1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHG--TMNSR 58
       ||||||||||||||||||||||||||||||||||||||||||||||||||||   ::::|
Db   1 MKRGKDDEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGNLSLSSR 60

Qy  59 SNNTSTLVHPGPSS-QPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQR 117
       ||||||||||||||  :|  ||||||||||||||::  ||:||||:|||||||:| ||||
Db  61 SNNTSTLVHPGPSNQQSCCGVERNLSVQHLDSSAAVHVTENFVSQMPFMENVMRSLAKHDQR 120

Qy 118 KMVREEDFAVPVVINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQN 177
       | :||||:||||.:|||||:||| |||||.|||||
Db 121 KTVREEDDFAVPVFVNSRRFQSHGSTKSGIVIEKHT------------------------ 156

Qy 178 VCLATCSKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSASSHDRVNDYNASLRQESR 237
        ||||||  :|||.|||:|.|||||  :|:: ||:|||||||||:||||| |:||||||
Db 157 -TLATCSK-LVRDKVKMNAKSGGFI--DLSSTEEVDLEKSASSYDRVNDCNSSLRQESR 211

Qy 238 NRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNREYSKSRACASLQQINEE 297
       |:||||||  ::|||:|  |||.|::||::||||||:|||:||:|  ||  |:||||||
Db 212 NKLYRDGGEAHMKDTANRVESHLVTESHSEEGHGSPDDDNGHEYCRSRGGVSLQQINEE 271

Qy 298 ASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKV 357
       |||||||:||||||||||||||||||.|||||||||||||:|||||||||||||||||||
Db 272 ASDDVSDNSMVDSISSIDVSPDDVVGALGQKRFWRARKAITNQQRVFAVQLFELHRLIKV 331

Qy 358 QKLIAASPDLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVV--KQRGDSEKTDQ 415
       |:|||||||::|:| :||||||||||||||| |||:|||||| |||  ||::||||||||
Db 332 QRLIAASPDIVLDEINYLGKVSAKSYPVKKLLVPSEFIVKPPLPQVVVNKQHRSDSEKTDQ 391

Qy 416 HKMESSAENVVGRLSNQGH--HQQSNYMPFANNPPASPAPNGYCFPQPPPSGNHQQWL 472
       ||||.||||||||||||||   :| ||||| :|||||||.||  :|||||||| |||
Db 392 HKMECSAENVVGRLSNQGHHHHNHQPSNYMPFPSNPPASPAVNGCCYPPQPPPSGN-QQWL 450

Qy 473 IPVMSPSEGLIYKPHPGMAHTGHYGGYYHMPTPMVMPQYHPGMGFPPPYGM 532
       |||||||||||||||||||.||||||||||  |:: ||||||||||||:||
Db 451 IPVMSPSEGLIYKPHPGMCHTGHYGGYYHGHYGGYYHGHMPPPMVMPFHPGMGFPPPMGNYFPPPYGV 510

Qy 533 MPTIMNPYCSSQQQQQ--QPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYP 591
       ||:    ||||||||   |||||||||||||||||||||||
Db    ...
```

Db 511 IPAMMNPYGPGQQQQQPQANEQTNQFGYSGNLQNNTHQES-SVNEAAPPQEPL-TKSYP 568
        || |||: |||||||.||||| |||| ||||||:||: :|||| |:||| :|||||
Qy 592 RARKSRQGSTGSSPSGPGPQGISGSKSFRPFAAVDEDSNINN--APEQTMTTTTTTTRTTVT 649
        |||||||:||.:||: .|:.|:|||:|||||:|||:.|:||: ||:|.|||||||||||||
Db 569 RARKSRQVSTASSASGREGISGSTSFRPFSAVDEDNNDNNNDAPDQMMTTTTTTTRTTVT 628

Qy 650 QTTRDGGGVTRVIK 663
        |||||||.||||||
Db 629 QTTRDGGEVTRVIK 642

RESULT 4
US-09-746-801A-61
; Sequence 61, Application US/09746801A
; Patent No. US2002008349A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 61
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Brassica sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (188)..(188)
; OTHER INFORMATION: Xaa = uncertain amino acid residue
; NAME/KEY: misc_feature
; LOCATION: (194)..(194)
; OTHER INFORMATION: Xaa = uncertain amino acid residue
; NAME/KEY: misc_feature
; LOCATION: (196)..(196)
; OTHER INFORMATION: Xaa = uncertain amino acid residue
; NAME/KEY: misc_feature
; LOCATION: (214)..(214)
; OTHER INFORMATION: Xaa = uncertain amino acid residue
US-09-746-801A-61

Query Match   24.6%; Score 904; DB 9; Length 324;
Best Local Similarity 48.4%; Pred. No. 1.3e-60;
Matches 218; Conservative 36; Mismatches 62; Indels 134; Gaps 16;

```
Qy  26 GGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSNNTSTLVHPGPSQPCGVERNLSVQH 85
        |||||||||||||| | : |:: ||.||::     ||:|:|  ||||:||||||||:|:|
Db   3 GGPRAPPRNKMALYEHLTTPSHRFTDHSS-SPRHTNTLFPPPPGPSNQPCGVERNLTSQH 61

Qy  86 LDSSAANQATEKFVSQMSFMENVRSSAQH--DQRKVMREEDFAVPVYINSRRSQSHGRT 143
        |||||  |    |  :|:|:|||::  | : ||||::|||:||||||::::
Db  62 LDSSASGH----VTQMSSMENVTTLAHRRGDQRKTLREEDDFAVPVYVND--------- 107

Qy 144 KSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVI 203
                                 :||||
Db 108 -------------------------SSRRFQ------------------------------ 113

Qy 204 SLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATE 263
                  |||||||: |:   ||  ||: : |:       |:|::||.||:|:||
Db 114 ----------CPLEKSASGCERV---NASCETESTSS--RLDHETGVMETDGVES----- 154

Qy 264 NHSQEGHGSPEDIDNREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVG 323
                  | : |:  |::       :||  |:||  ::|:||||||||: |:||||
Db 155 ----------HGNPNDVDDD-------DDD--DSISSIDVSSDEVVG 182

Qy 324 ILGQKRFWRARKAIA-NQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKS 382
        :||| ||||| | : ||||:||||||||||:|||:|:|||:|:|:||||||:||.
Db 183 VLGQNXFWRARXAXAKNQQRIFAVQLFELHRXIKVQRLIASSSDVLLDEISYLGNV--- 238

Qy 383 YPVKKLLPSEFLVK-PPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQG-HHQQSNY 440
         ||||||||| |:| ||||||| :|:.||||||||:|||||||||||: |||| ||| ||||
Db 239 -PVKKLLSEFIILKPPPLPQVTKHRSSDSEKTDQNKMESSAENVVGKSSNQGQQHQPSNY 297
```

```
Qy  441  MPFANNPPASPAPNGYCFPPQ-PPPSGNFQ  469
              |||||:|||||   :||  |||||||||
Db  298  MPFASNPPAA---NGCYPPQHPPPSGGNQ  324


RESULT 5
US-09-746-801A-59
; Sequence 59, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59
; LENGTH: 599
; TYPE: PRT
; ORGANISM: Lycopersicum esculentum
US-09-746-801A-59

Query Match       21.2%;  Score 777;  DB 9;  Pred. No. 1.5e-50;
Best Local Similarity   33.4%;
Matches  232;  Conservative  95;  Mismatches 216;  Indels 152;  Gaps 29;

Qy   45  PSQRFGDHGTMNSRNNTSTLVHPGPSSQPCGVERN-LSVQHLDSSAANQATEKFVSQM  102
Db    1  PSQRF-NSGVLPLDPNNTSKMAPPS-SSQGSGHDRSGYLPIQHPPSRRLADKPPGHSSDP  58

Qy  103  SFMENVRSSAQHDQRKMVREEDFAVPVYINSRRSQSHGRTKSGIEKEKHTP---MVAPS  159
Db   59  STL------LQQYELKKRTEEDDFTVPIFVNSKLGQAHG--SHNVNMEKLSPSGQLFCPN  110

Qy  160  ------SHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDV---SVT  210
Db  111  KELEGVTHLTLRQQRNSQ--NKENL--KCTLAR-REKTTSNSASKECRLDPQVGCSSIP  164

Qy  211  EEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGH  270
Db  165  EPVKGTYDGSSYPRKEFVSEQL  186

Qy  271  GSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRF  330
Db  187  -TANDLVNDTESQEDRAHKSLQTGNLDRGDDLSETSRVESISGTDISPDDIVGIIGLKRF  245

Qy  331  WRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLP  390
Db  246  WKARRAIVNQQRVFAIQVFELHRLIKVQRLIAGSPNSSLEDPAYLGK-PLKSSSIKR-LP  303

Qy  391  SEFIVKPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLS---NQGHHQQSNYMPFANNP  447
Db  304  LDCIVRE--SQSVLKRKHDSEK-PHFRMEHTAESNVGKASLSTVQNGSQLSSHKPFSGTP  360

Qy  448  PASPAPNG------YCFPPQPPPSGNHQQWLIPVMSPSEGLIYKPHGMAHTG----  494
Db  361  LPPTPVTNDSNAGPWCF-QQPPG--HQWLIPVMSPSEGLVYKFPFPGPGFTSPICGSGPP  415

Qy  495  --HYGGYYGHYMPTPMVMPQYHPGMGF-------PPPGNGYFPPYGMGMPTIMNPYCSSQQQQ  547
Db  416  GSSPTMGNFFAPTYGVPAPNPHYQ-GMGVPAPPTGHGYFRQYGM--PAMNPPISS----  468

Qy  548  QQQPNEQMNQFGHPG--------------NLQNTQQQ--QQRSDNEPAPQQQQQPTKSYPR  592
Db  469  --TASEESNQYTMPGLQHQFSGVVDDVNIQHQDSSNVLNQKKENVPDDVVRYQSTKDNEVQ  526

Qy  593  ARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTT  652
Db  527  A------SSASSPIETAGRNMLSLFPTISPVTDNRDGSPQACVPDNP-----------  566

Qy  653  RDGGGVTRVIKVVPHNAKLASENAARIFQSIQEER  687
```

```
Db  567  ------ARVIKVVPHNARSATESVARIFQSIQQER  595


RESULT 6
US-09-746-801A-27
; Sequence 27, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Oryza sativa
US-09-746-801A-27

Query Match       21.1%;  Score 773.5;  DB 9;  Pred. No. 3.9e-50;
Best Local Similarity   32.1%;
Matches  261;  Conservative 105;  Mismatches 257;  Indels 191;  Gaps 36;

Qy    4  GKDEE-KILEPMFPRLHVNDADK-GGPRAPPRNKMALYEQLSIPSQRF-----GDHGTMNS  57
Db   12  GKEAKGKVMGPLFPRLHVNDAAKGGGPRAPPRNKMALYEQFTVPSHRFSGGGGGGGVGGS  71

Qy   58  RSNNTSTLVHPGPSSQPCGVERNL------------------SVQHLDSSAANQATEKFVSQMSF  104
Db   72  PAHSTSAASQSQSQSQVYGRDSSLFQPFNVPSNRPGHSTEKINSDKINKKISSGRKELGM  131

Qy  105  MEN-----------VRSSAQHDQR------------KMVREEDFAVPVYINSRRSQSHGRTK  144
Db  132  LSSQTKGMDIYASRSTAEAPQRRAENTIKSSGKRLADDDEFMVPSVFNSRFPQYSTQEN  191

Qy  145  SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV-CLATCSKPEVRDQVK-ANARSGGFV  202
Db  192  AGVQ-DQSTPLVAANPHKS------PSTVSKSSTKCYNTVSKKLERIHVSDVKSRTPLKD  244

Qy  203  ISLDVSVT-EEIDLEKSASSHDRVNDYNASLRQESRN-RLYRDGGKTR-LKDTD--NGA  256
Db  245  KEMEAAQTSKNVVEVEKSSSFHASKDMF-------ESRHAKVYPKMDKTGIINDSDEPHGGN  298

Qy  257  ESHLATENHS------------QBGHGSPEDIDNDREY------SKSRACASLQQIN  295
Db  299  SGHQATSRNGGSMKFQNPPMRRNEISSNPSSENTDRHYNLPQGGIEETGTKRKRLLEQHD  358

Qy  296  EEASDDVS--------------DDSMVDSISSIDVSPDDVVGIILGQKRFWRARKAIANQQ  341
Db  359  AEKSDDVSRLLEQHDAENIDDVSDSSVECITGWEISPDKIVGAIGTKHFWKARRAIMNQQ  418

Qy  342  RVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKLLPSEFLVKPPLPH  401
Db  419  RVFAVQVFELHKLVKVQKLIAASPHVLIESDPCLGNALLGS--KNKLVEENLKAQPLLV  475

Qy  402  VVVKQRGDSEKTDQHKMESSAENV-------VGRLSNQGHHQQSNYMPFAN---NPPA  449
Db  476  ATI------DDVEPSLQQPEVSKENTEDSPPSPHD'TGLGSGQRDQAATNGVSKSNRRATPVA  531

Qy  450  SP-APNGYCFPPQPPSGNHQQWLIPVMSPSEGLIYKPHG---------------M  490
Db  532  SDNKQNNWGVQLQPP--QNQWLVPVMSPLEGLVYKPYSGPCPPAGSILAPFYANCTPL  587

Qy  491  AHTGHYGGYYGHYMPTMVMPQYHPGMGFPP-PGNGYFPPYGMGMPTIMNPYCSSQQQQ  548
Db  588  SLPSTAGDFMNSAYGVPMPHQPHMGAPGPPSMPMN-YFPPFSI-PVMNP--------  635

Qy  549  QQPNEQMNQFGHPGNLQ---NTQQQQRSDNEPAPQQQQQPTKSYP---RARKSR--  597
Db  636  TAPAPVVEQGRHPSMPQPYGNFEQQSWISCN------MSHPSGIWRPHASRDSEA  684
```

```
Qy   598 QGSTGSSPSPGQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVTQTTRDGGG 657
Db   685 QASSASSPFDRFQCSGS---GPVSAPFTVSAQNNPQP------------SYSSRD--N 726
Qy   658 VTRVIKVVPHNAKLASENAARIFQSIOEERKRYD 691
Db   727 QTNVIKVVPHNSRTASESAARIFRSIQMERQRDD 760
```

RESULT 7
US-10-437-963-136891
; Sequence 136891, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 136891
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_38427C.1.pep
US-10-437-963-136891

Query Match   21.1%;   Score 773.5;   DB 16;   Length 760;
Best Local Similarity 32.1%;   Pred. No. 3.9e-50;
Matches 261; Conservative 105; Mismatches 257; Indels 191; Gaps 36;

```
Qy     4 GKDEE-KILEPMFPRLHVNDADK-GGPRAPPRNKMALYEQLSIPSQRF----GDHGTMNS 57
Db    12 GKEAKGKVMGPLFPRLHVNDAAKGGGPRAPRNKMALYSQFTVPSHRFGGGGGGGGVGGS 71
Qy    58 RSNNTSTLVHPGPSSQPCGVERNL------------SVQHLDSSAANQATEKFVSQMSF 104
Db    72 PAHSTSAASQSQSQSQVYGRDSSLFQPFNVPSNRPGHSTEKINSDKINKKISSGRKELGM 131
Qy   105 MEN--------VRSSAQHDQR---------KMVREEEDFAVPVYINSRRSQSHGRTK 144
Db   132 LSSQTKGMDIYASRSTAEAPQRRAENTIKSSSGKRLADDDEFMVPSVFNSRFPQYSTQEN 191
Qy   145 SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV-CLATCSKPEVRDQVK-ANARSGGFV 202
Db   192 AGVQ-DQSTPLVAANPHKS-------PSTVSKSSTKCYNTVSKKLERIHVSDVKSRTPLKD 244
Qy   203 ISLDVSVT-EEIDLEKSASSHDRVNDYNASLRQESRN-RLYRDGGKTR-LKDTD---NGA 256
Db   245 KEMEAAQTSKNVEVEKSSSFHASKDMF------ESRHAKVYPKMDKTGIINDSDEPHGGN 298
Qy   257 ESHLATENHS--------------QEGHGSPEDIDNDREY------------SKSRACASLQQIN 295
Db   299 SGHQATSRNGGSMKFQNPPMRRNEISSNPSSENTDRHYNLPQGGIEETGTKRKRLLEQHD 358
Qy   296 EEASDDVS------------DDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQ 341
Db   359 AEKSDDVSRLLEQHDAENIDDVSDSSVECITGWEISPDKIVGAIGTKHFWKARRAIMNQQ 418
Qy   342 RVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPH 401
Db   419 RVFAVQVFELHKLVKVQKLIAASPHVLIESDPCLGNALLGS---KNKLVEERNLKAQPLLV 475
Qy   402 VVVKQRGDSEKTDQHKMESSAENV-----------VGRLSNQGHHQQSNYMPFAN---NPPA 449
Db   476 ATI----DDVEPSLQQPEVSKENTEDSPSPHDTGLGSGQRDQAATNGVSKSNRRATPVA 531
Qy   450 SP-APNGYCFPQPQPPPSGNHQQWLIPVMSPSEGLIYKPHPG---------M 490
Db   532 SDNKQNNWGVQLQPP----QNQWLVPVMSPLEGLVYKPYSGPCPPAGSILAPFYANCTPL 587
Qy   491 AHTGHYGGYYGHYMPTPMVMPQYHPGMGFPP--PGNGYFPPYGMGMPTIMNPYCSSQQQQQ 548
Db   588 SLPSTAGDFMNSAYGVPMPHQPQHMGAPGPPSMPMN-YFPPFSI--PVMNP--------635
Qy   549 QQPNEQMNQFGHPGNLQ--NTQQQQRSDNEPAPQQQQQPTKSYP---RARKSR-------597
Db   636 TAPAPVVEQGRHPSMPQPYGNFEQQSWISCN------------MSHPSGIWRFHASRDSEA 684
Qy   598 QGSTGSSPSPGQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVTQTTRDGGG 657
Db   685 QASSASSPFDRFQCSGS---GPVSAPFTVSAQNNPQP------------SYSSRD--N 726
Qy   658 VTRVIKVVPHNAKLASENAARIFQSIOEERKRYD 691
Db   727 QTNVIKVVPHNSRTASESAARIFRSIQMERQRDD 760
```

RESULT 8
US-10-424-599-230456
; Sequence 230456, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 230456
; LENGTH: 534
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_50123C.1.pep
US-10-424-599-230456

Query Match   19.8%;   Score 725.5;   DB 12;   Length 534;
Best Local Similarity 40.0%;   Pred. No. 1.1e-46;
Matches 199; Conservative 64; Mismatches 157; Indels 77; Gaps 21;

```
Qy   235 ESRNRLYRDGG--KTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQ 292
Db    75 ESRSNDIGDNGCLVESARETDKG---NAPTANQI----SPADAINDTEHHDTRMGSPIQ 126
Qy   293 QINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELH 352
Db   127 RGNLNESDNASKISMVENLSTVRISPDDVVGIIGQKHFWKARRAIANQQRVFAVQVFELH 186
Qy   353 RLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEK 412
Db   187 RLIKVQQLIAGSPDILLEDGAFLGKSPPKGSTPKK--LALEYVVKPRQQN--LKRKDDSEK 243
Qy   413 TDQHKMESSAENVVGRLS---NQGHHQQSNYMPFANNP----PASPAPNGYCFPPQPP 463
Db   244 LN-HKMECSAENAVGKTSLSSVKDGSH-LSKCTPFGNQHQTNVAADSGMGPWCFNQSPP 301
Qy   464 PSGNHQQWLIPVMSPSEGLIYKPHPGMAHTG-HYGG------YYGHYMPTPMVMPQY 513
Db   302 ---GHPWLIPVMTPSEGLVYKPYPGPGFTGTGCGGGGGTLREDDRRGSFMNPGYGIPTS 357
Qy   514 HPGMGFPP----PG-NGYFPPYGMGMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQ 569
```

```
Db   352 DNAEKIDDLSDSS-VECITAWEISPDEIVGAIGAKHFWKARRAIINISLMKAKYNDILIL 410
                 |||   ||     ||  |        |        | ||  |:|
Qy   340 --QQRVFAVQLFELHRLIK---------------------------VQKLIAASPDLLLDEISF 374
                 ||||||:||||||:                        |:| :
Db   411 LYQQRVFAAQVFELHKLVKVSSWIEFGVEYENNQNLYLTYKQVQKLIAASPHVLIEGDPC 470
             ||||||  |||||:|                         |||||||    || |::
Qy   375 LGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQH--KMESSAENVVGRLSN 431
             ||    |:| |    |:|  |  |  |   ::|:|           |   ||||  :::
Db   471 LGNALLAS---KKKMAEENLKAQP--VLVATNDVQPSLQEPELSKENSENPPSPRDT 524
             ||||||:            ||||       |:|||||||||||:| ||:::::
Qy   432 ---QGHHQQSNYMPPANN------PPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLI 483
                :|||||  ||             |  |    |       |   |||||||||||||
Db   525 APVSGHHDQTAKIGASKSNLRATFVASDNRQNNCGVQLQPP--QNQWLIPVMSPSEGLV 581
                     |      ||   |      |  |:|:||||||||||
Qy   484 YKPHPG-------------------MAHTGHY-GGYYGHYMPPTMVMPQYHPGMGFPP 521
             :|::                      ::  |   |||||||||||||  |||
Db   582 YKPYSGPCPPAGSILAPFYANCTPLRLPSTTGDFMNSAYG--VPIPH-QPQHMGAPGTPT 638
             ||| ||| | | || |    |  ||  ||  | ||:||       ||     |  ||||
Qy   522 PGNGYFPPYGMMPTIMNPYCSSQQQQQOPNEQMNQFGHPGNLQ---NTQQQQQRSDNEP 578
             | |||| ||  ||  | |:||||||         |  ||            ||:|  |
Db   639 MPMNYFPPFSV--PVMNPVALASAVFQGR-----------HPSMPQPYGNLEQHSRMSCN-- 685
                 |  |      |  ||| | |  |               |  | |    ||
Qy   579 APQQQQQPTKSYP----RARKSR---QGSTGSSPSGPQGISGSKSFRPFAAVDEDSNIN 630
             |           |    |  |:|     | :  |||||   |  ||
Db   686 ----------MSHPSGIWRFHASRDSEAQASSASSPFDRLQCGGSGPVSAFP------- 727
                       :|  | |  |  | |  |  |    |||    |  |
Qy   631 NAPEQTMTTTTTTRTTVTQTRDGGVTRVIKVVPH-NAKLASENAARIFQSIQEERKR 689
                    |::    |:    ||   |||||||  | | | | | :| | ||||||
Db   728 --------TASAQNTQQPSSGSRD--NQTNVIRVIPHNSQTASESAARIFRSIQMERQQ 778
                                |  |       ||  | | | |:| | | ||||:||:|||
Qy   690 YDS 692
             ||
Db   779 DDS 781
```

RESULT 10
US-09-746-801A-23
; Sequence 23, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 23
; LENGTH: 389
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-746-801A-23

```
Query Match            16.9%;    Score 621.5;  DB 9;    Length 389;
Best Local Similarity  39.0%;    Pred. No. 6.3e-39;
Matches    164;  Conservative    58;  Mismatches   132;  Indels    67;  Gaps   17;

Qy   297 EASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIK 356
                 |  | | |  || :  |||  || | |  | | ||||  |  ||| || |||||||
Db     2 DRGDDLSETSRVESISGTDISPDDIVGIIGLKRFWKARRAIVNQQRVFAIQVFELHRLIK 61
         |||| ::  | |  :  ||  ||  ||:||  |||||||  |  |||| |: |||||||
Qy   357 VQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQH 416
             |||||||||||||||| :|  |||  |    || | | ||  | |: | :  |  | |
Db    62 VQRLLAGSPNSSLEDPAYLGK-PLKSSSIKR-LPLDCIVRE--SQSVLRKKHDSEK-PHF 116
             ||| || | |||| | ||  | |    |      |   |      |   |  |  |
Qy   417 KMESSAENVVGRLS--NQGHHQQSNYMPPANNPPASPAPNG-----YCPPPQPPPSGNH 468
             |||||||||| |    || |||||  |  || ||| | |       | ||| || ||
Db   117 RMEHTAESNVGKASLSTVQNGSQLSSHKPFSGTPLPTPVTNDSNAGPWCF---QQPSGH- 172
              | |||||| |   |  ||| |  ||  ||  |  |  |  ||            || ||
Qy   469 QQWLIPVMSPSEGLIYKPHPGMAHTG----------------HYGCYYGHYMPTPMVMPQYHPG 516
             |||||||||||||| ||||                     |  | | | ||| | |  |
Db   173 -QWLIPVMSPSEGLVKPFSGPGFTSPICGSGPSGSSPTMGNFFAPTYGVAPNPHYQ-G 230
              ||||||||||||: ||  ||   |    | |  | |  |  |   | | || |  |
```

RESULT 9
US-10-437-963-131862
; Sequence 131862, Application US/10437963
; Publication No. US20040123343A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; APPLICANT: Wu, Wei
; APPLICANT: Boukharov, Andrey A.
; APPLICANT: Barbazuk, Brad
; APPLICANT: Li, Ping
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53221)B
; CURRENT APPLICATION NUMBER: US/10/437,963
; CURRENT FILING DATE: 2003-05-14
; NUMBER OF SEQ ID NOS: 204966
; SEQ ID NO 131862
; LENGTH: 781
; TYPE: PRT
; ORGANISM: Oryza sativa
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT4530_33888C.1.pep
US-10-437-963-131862

```
Query Match            18.7%;    Score 686.5;  DB 16;   Length 781;
Best Local Similarity  28.9%;    Pred. No. 1.8e-43;
Matches    263;  Conservative   113;  Mismatches   223;  Indels   39;  Gaps   39;

Qy     4 GKDEE---KILEPMFPRLHVNDADK-GGPRAPPRNKMALYEQLSIPSQRF-GDHGTMNSR 58
             |  |   |  |   |:  |  | | |||| | |||||| | || || |   | | | |
Db     8 GKEVEERGKVMGPLFPRLHVNDAAKGGGPRAPPRNKMALYEQFTVPSHRFSGGGGALASA 67
             ||  | |: || ||||||||| ||| |||||||||||||:   | || |
Qy    59 SNNTSTLVHPGPSSQPCVERNL--------------------SVQHLDSSAAN--------- 92
             |  | | | | ||  |  |                     || | || ||
Db    68 RGSLARSTSAASQSQVYGCDMPLFPEFNVPSNGPGQSVEKMNSNSVNRQINGSRKDSGML 127
               |     |:|| |  ||     || ||   | |||  |  | |  |  || |   ||
Qy    93 ----QATEKFVS---QMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRSQSHGRTK 144
                 ||   |     |  | | | ||  |  |  |  ||| |  | | |  |  ||
Db   128 STQPKGIDKYGSGSRAECAPQQRVEKGIKSSSGRKLADDDEFTVPSVFSARFPQYSTKER 187
             | |    ||  | |   | |     |  |   |  |  ||| | | |   |  ||
Qy   145 SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSK------PEVRDQVKANAR 197
             |  |  |  || | |  || ||  |   |  | |           |  ||| || |
Db   188 AGVGEES-TPLVALSPHKSP--PAVSKSPTK---CYNTVSKNLERINVSDVKSRGSQKDK 241
             || | |  |||| ||| ||   |   ||       | | |  | |  |  |  | |
Qy   198 SGGFVISL-DVSVTEEIDLEKS-----ASSHDRV------NDYN-ASLRQESRNRLYRDG 244
             |||  |:|  |||  ||  ||       |  | |       | |  | |  |  | | |
Db   242 ETGPAQTLKNVEVEHFSSFEASKDMFGSKHAKVCPKTGTINDLLDEPHLENSEHQATSRNG 301
             | |    |  |  |  | |  |  ||   |   | |  | |||  | | || |  || ||
Qy   245 GKTRLKD---------TDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQ 293
             |  | ||          |  | |   |   | | | || |  |  |  | |    ||
Db   302 SSVKFQNPFVRRNTISAKPSPGIEN---TNGHCNLPQGGLKEAGTKRKRLEA------Q 351
             |  | |  | | |  || |||        || | |||  | |  || | ||
Qy   294 INEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIAN------------- 339
              |  |  ||| | |  ||  ::| | | | | |||  ||||| |
```

```
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 57
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Zea mays
US-09-746-801A-57

Query Match    12.5%;  Score 459.5;  DB 9;  Length 317;
Best Local Similarity   32.1%;  Pred. No. 1.1e-26;
Matches 136;  Conservative 51;  Mismatches 98;  Indels 139;  Gaps 15;

Qy  297 EASDDVSDDSMVDSISSIDVYSPDDVVGILGQKRFWRARKAIANQRVFAVQLFELHRLIK  356
            ||| ::|: ||||  :|:|| |||||||  || ||| ||:|.||||||:|||||||||
Db    3 EQNDDLSDSS-VESLPGMEISPDDVVSAIGPHFWKARRAIVNQQRVFAVQVFELHRLIK   61

Qy  357 VQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLIVKPPLPHVVVKQRGDSEKTDQ-  415
            ||||||||     |:  ||.  |:      :  |: |:
Db   62 VQKLIAASPHVLIEGDPCLGKSLAVS--KKRLAGD------VETQLESAKNDDG  107

Qy  416 --HKMESSAENVVGRLSNQGHHQQSNYMPPANN------PPASPAPNGYCFPFQPQPP  464
               :  :       :  | :. :||
Db  108 VRPTQLEHSKEKTEANQPSPSQDEQA----ATNGDVAALMHTPSDNKQKSWCIP--APP  160

Qy  465 SGNHQQWLIPVMSPSEGLIYKPHFGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGN  524
                 |||||||||||:||      :|      |  ||:
Db  161 S-----QWLIPVMSPSEGLIVKP-----YTGH------CPPVGS  188

Qy  525 GYFPPYGMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQ  584
            |:
Db  189 LLAPPF------  194

Qy  585 QPTKSYPRARKSRQGST-GSSPSGPQGISGSKSFRPF------AAVDEDSNI  629
                ||      .:  :   ||  |     :      |: :|
Db  195 --FASYPTSSSSTAGGDFMSSACGARLMSAPVYFPSFSMPAVSGSAVEQVSHVAASQHKR  252

Qy  630 NNAPEQTMTTTTTTTR--TTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEER  687
             |.  :          : |:    |     |:||||||||.| ||||:| ||||:||:||
Db  253 NSCSEAVLASRDSEVQGSSASSPASSETAAQPRVIRVVPHTARTASESAARIFRSIQMER  312

Qy  688 KRYD  691
        |: |
Db  313 KQND  316


RESULT 13
US-09-746-801A-33
; Sequence 33, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 33
; LENGTH: 540
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-746-801A-33

Query Match    11.9%;  Score 436.5;  DB 9;  Length 540;
Best Local Similarity   23.6%;  Pred. No. 1.4e-24;
Matches 177;  Conservative 98;  Mismatches 195;  Indels 279;  Gaps 32;

Qy    5 KDEEK-ILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLS------IPSQRFGDHGTMN  56
            |||:|  | |::||   |.  |:             :: : :|
Db    5 KDEAKRTTIPPLFPRVHVNDTGRGG------LSQQFDGKTMSLVSSKRPNLPSPTN  54

Qy   57 SRSNNTSTI---VHPGPSSQPC--GVERNLSVQHLDSSAANQATEKFVS---QMSFMEN  107
            :::   :  |:      :|  :   :     :  :       ::  :
```

```
Qy  517 MGF---PPPGNGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQN-----T  567
            ||    |  |||:|||||  |  ||:                :  :
Db  231 MGVPFAPPTCHGYFRQYGM--PAMNPPISS------TASEESNQYTMPG-LQHQFSGVVD  281

Qy  568 QQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGST-GSSPSGPQGISGSKSFRPFAAVDED  626
                 :    :
Db  282 DVQHSTSGLSNVLNQKKENVPDVVRYQSTKDNEVQASSASSPIETAGRNMLSLFPTSPVT  341

Qy  627 SNINNAPEQTMTTTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEE  686
             ::|:          |||:||||||||| |       ||:||||  |||: ||||:|||:
Db  342 DNRDGSPQACVPDNP--------ARVIKVVPHNARSATESVARIFQSIQQE  384

Qy  687 R  687
        |
Db  385 R  385


RESULT 11
US-09-746-801A-63
; Sequence 63, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 63
; LENGTH: 168
; TYPE: PRT
; ORGANISM: Brassica sp.
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (43)..(43)
; OTHER INFORMATION: Xaa = uncertain amino acid residue
; NAME/KEY: misc_feature
; LOCATION: (65)..(65)
; OTHER INFORMATION: Xaa = uncertain amino acid residue
US-09-746-801A-63

Query Match    13.5%;  Score 495.5;  DB 9;  Length 168;
Best Local Similarity   59.9%;  Pred. No. 8.3e-30;
Matches 115;  Conservative 11;  Mismatches 11;  Indels 29;  Gaps 11;

Qy  486 PHPGMAHTGHY-GGYYGHYMPTPMVM------PQYHPGMGFPPPGNGYFPPYGMMPTIM  537
            | | |:||.  ||||||:|:|:|| |      | |||||||  ||||||||||  ||:
Db    1 PGPDPGHTGPVCGGYYGHMPAPMFMGGGGGGQPPFHPGMGFXSHGNGYFPPYGGI--MM  58

Qy  538 NPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSYPRARKSR  597
            |||:   ||||||||||:     ::    |: : |:      |||||||||||:|||
Db   59 NPYYSG-XQQQQQPNEQMN-----NNI----QQQSSVNE-ATSQQQQPTKSYPRAKKSR  106

Qy  598 QGSTGSSPSGPQGISG-SKSFRPFPAAVDE-DSNINNA--PEQTMTTTTTTTTRTTVTQTT  652
            |            : | |||| ||  ||:   |:|||| || |:||||||||||||||||
Db  107 Q------EGISGKKKSFQPFSAVDDVHDDKINNAAQPTEEMTTTTTTTTTTTVTQTT  156

Qy  653 RDGGGVTRVIKV  664
        |||.|||||||
Db  157 RDGAGVTRVIKV  168


RESULT 12
US-09-746-801A-57
; Sequence 57, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
```

```
Db    55 NISDSLSTFSLSLPPPNNARLIDGPEKNQFSPIYNTKFEGKLNKKGINYTSPKGSSVTN 114

Qy   108 VRSSA--QHDQRKMVREEBDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIR 165
               : :  ||| : ::  :|| :                       |::
Db   115 TKPSSIKQNEYLKNLTSLDSIKSPIVIHSE------------IDPQANTDL-----SLQ 156

Qy   166 FQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRV 225
                                              ||:|
Db   157 F-------------CTSGGSSKP--------GGEAVVGSKILLSE----------RL 181

Qy   226 NDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKS 285
               ::  :|  :::  :||: :| :      :  || : : :
Db   182 EDEN--------------QNGSPNVMKTQSYRR----NFABFNNETQ-KKP 213

Qy   286 RACASLQQINEBEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFA 345
               :  :|  ::||||||| || ||||||||||||||||||||||||||||||
Db   214 KTLPRREQV------ASNCSAIESLSGISASSYDIARVIGEKRFWKMRTYMINQQKIFA 266

Qy   346 VQLFELHRLIKVQLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVK 405
               ||||||||| ||||||||||||||||||||||||||||||||||||||||||||||
Db   267 GQVFELHRLIMVQKMVAKSPNLFLES-------------------K 293

Qy   406 QRGDSEKT--DQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGY------ 456
               ||:   ::| :|| || ||||| ||||||||||||||||||
Db   294 LNGVKHGTMRSSHQLAMAASKV--RKPNTENHK-------------PVPEEYPEHMKPK 337

Qy   457 ------CFPPQ-PPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGH 502
                         : |||||||||||||||||||||||||
Db   338 LPLPSISKELVTPINFQQLLPPPGN--QWLVPVITDSDGLVYKPFPG------ 382

Qy   503 YMPTP------MVMPQY---HPGMGFP--PPGNGYFPPYGMMPTIMNPYCSSQQQQQQ 549
               |::||       ||||||                  ||:
Db   383 --PCPPSSSAFMVPVYGQDSLETPPRFPVSSSPFSHSYFPPPNARTTV--------- 427

Qy   550 QPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQP---TKSYPRARKSRQGSTGSSPS 606
               |||| ||| |||||||||||              ||||||||||||||
Db   428 ---DQTNPFG---------QFQRWNTSSHMTQAIPFSLKKSQESNDSDIHGSTASSP- 473

Qy   607 GPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTTRTTVTQTT---RDGGGVTRVI 662
               |||||||||||||||||||||||||||||||
Db   474 ------------------------------PEKHKLEVLPLFPTEPTHQTDEYKQQPMLRAI 507

Qy   663 KVVPHNAKLASENAARIFQSIQEERKRYD 691
               |||||||||||||||||||||||||||
Db   508 KAVPHNSTSASESAARIFRFIQEERRDSD 536
```

```
RESULT 14
US-09-746-801A-29
; Sequence 29, Application US/09746801A
; Patent No. US2002008349A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; SOFTWARE: PatentIn version 3.1
; NUMBER OF SEQ ID NOS: 68
SEQ ID NO 29
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (88)..(88)
; OTHER INFORMATION:
; NAME/KEY: unsure
; LOCATION: (98)..(98)
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
; NAME/KEY: unsure
; LOCATION: (113)..(113)
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
```

```
; NAME/KEY: unsure
; LOCATION: (114)..(114)
; OTHER INFORMATION: amino acid not confirmed, based on nucleic acid sequence
US-09-746-801A-29

Query Match             10.8%;  Score 397;  DB 9;  Length 248;
Best Local Similarity   29.5%;  Pred. No. 4.8e-22;
Matches 121;  Conservative 34;  Mismatches 75;  Indels 180;  Gaps 13;
```

```
Qy   297 EASDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIK 356
               : ||| ::| ::|||||| ||||||||||||||||||||||||||||||||||||
Db     3 EQNDLSDSS-VESLPGMEISPDDVVSAIGPKHFWKARRAIVNQQRVFAVVFEELHRLIK 61

Qy   357 VQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQH 416
               |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    62 VQKLIAASPHVLIEGDPCLGKSLAVSXK--------------------- 89

Qy   417 KMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVM 476
                |||||||||||||||||||||||||||||||||||||||| |   |: ||||| |:
Db    90 ------------RLS-------------------------------QWLLPXM 99

Qy   477 SPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHP--------GMGFPPPGNG--- 525
               |||||:|:        :  | |:                       :
Db   100 SPFEGLVYKPYPGXXPSG-----GSLLAPPFAS--YPTSSSTAGGDFMSSACGARLM 151

Qy   526 ----YFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQ 581
                            |  ::|       ::
Db   152 SAPVYFPSFS-MPAV---------------SGSAVEQVSHVA----ASQHKRNSCSEAVLA 192

Qy   582 QQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTT 641
               :  :| :
Db   193 SRDSEV------------QGSSASSPASSE------------------------ 210

Qy   642 TTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYD 691
                                  |  ||| :|| :: : :||:|| : |
Db   211 -----TAAQ-------PRVIRVVPHTARTASESAARIFRSIQMERKQND 247
```

```
RESULT 15
US-10-424-599-222485
; Sequence 222485, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
SEQ ID NO 222485
; LENGTH: 335
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_42933C.1.pep
US-10-424-599-222485

Query Match             10.7%;  Score 393.5;  DB 12;  Length 335;
Best Local Similarity   34.0%;  Pred. No. 1.4e-21;
Matches 127;  Conservative 49;  Mismatches 116;  Indels 81;  Gaps 19;
```

```
Qy   358 QKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHK 417
               ||||||||||||||||: ||:|
Db     3 RKIIAASPDVLLEDVAFMGKPPLTESPPKN-ISLEVVVEPQQQN--PKRKNDSEKLN-HK 58

Qy   418 MESSAENVVGRLSN-----QGHHQQSNYMPFANNP-PASPAPNG----YCFPPQPPPSGN 467
               ||| | ||||| |         ::| |  | :  |: |  :
Db    59 TECSARNAVAKRTSFSSPKNGSH-LANHTPFSGTPHQANVASDNKTSPWCFNQTP----- 112
```

```
Qy   468  HQQWLIPVMSPSEGLIYKPH--PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPP--  522
          ||||||||||||||||||||  ::   ||  | ||  ::
Db   113  GHQWLIPVMSPSEWLYYKPYPQPGFMRTMH-GGWHKPLVLAPLSATFMNPVVQFPATHAL  171

Qy   523  --------GNGYFPYGMMPTIMNPYCSSQQQQQPNEQMNQFGHPGNL-QN--------  566
                   ||||| ::   :|    :
Db   172  VRVSPFVPELSHTYFALFGM--RVVN------QATSGSAVEQVNQFAAQGSHGQNGHSSV  223

Qy   567  -TQQQQQRSDNEPAPQQQQQPTKSYPRARKSR--QGSTGSPSPSG-PQGISGSKSFR  618
                    ||||||| ||||||||||||  || |||||  ||||||
Db   224  EGADFNTHHNQSSSNLPVQKNGARLHVKKSQALKERGLQGSTRSSPSEMAQGIRAGKI-  281

Qy   619  PFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAAR  678
               :  :    ::|                  |:|:|  |||||
Db   282  --------ADGSDAQSLSLHAVETRQQ-----------TQAIKVVPHNRKSATESAAR  320

Qy   679  IFQSIQEERKRYD  691
          |||||||||::|
Db   321  IFQSIQEERKQHD  333
```

```
Query Match        10.3%;  Score 377;  DB 12;  Length 334;
Best Local Similarity  33.4%;  Pred. No. 2.5e-20;
Matches  124;  Conservative  45;  Mismatches  108;  Indels  94;  Gaps  16;

Qy   192  VKANARSGGFVISLDVSVTEEIDLE--KSASSHDRVNDYNASLRQESRNRLYRDGGKTRL  249
          ||:                     ||  ::|
Db   25   VKSKGPSG---IKEKEPVQVRIDLEDKETTPSFQVLNDKTWSPDPKLSSHM------DRL  75

Qy   250  KDTDNGAESH-LATENHSQEGHGSP------------------EDIDND----REYSKS  285
          || |:::     || ::|                              :  :|
Db   76   KKQHAEAESYQIRTRNENAVETQSPPKNGVSLLSKPYVDRREQNGDSDLLGHLRETGVK  135

Qy   286  RACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFA  345
             | ||:|||::||||||||||||||||||||||||||||||||| ||||:|||||||||
Db   136  R---KRSHHDVEQNDDLSDSS--VESLPGMEISPDDVVSAIGPKGHFWKARRAIVNQQRVFA  191

Qy   346  VQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPLPHVVVK  405
          |||||||||||||||||||              :  : |  | |:        |:
Db   192  VQVFELHRLIKVQKLIAASPHVLIEGDPCLGKSLAVS--KKRLIAGD----------VE  237

Qy   406  QRGDSEKTDQ----HKMESSAENVVGRLSNQGHHQQ---SNYMPFANNPPASPAPNGYCF  458
          ||   :  | :     :                  :|
Db   238  TQLESAKNDDGVRPTQLEHSKERTEANQPSPSQDEQAATNGDVAASMHTPSDNKQKSWCI  297

Qy   459  PPQPPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYCHYMPTPMVMPQYHPGMG  518
          :          :||||||||||::
Db   298  P--APPS---QWLIPVMSPSEGLVYKP-----YTGH-----------------------  323
```

```
Qy   519  FPPPGNGYFPP  529
               ||||  ||
Db   324  CPPVGSLLAPP  334
```

```
Query Match         8.6%;  Score 316;  DB 12;  Pred. No. 1.4e-15;
Best Local Similarity  28.4%;  Length 383;
Matches  118;  Conservative  51;  Mismatches  124;  Indels  122;  Gaps  18;

Qy   1    MKRG--KDE--EKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTM  55
          ||:   ||   |:|||||:|||||:||||||||||||||||||||||:||:|:
Db   11   MRRGATKDDAAPDKVMGPLFPRLHVNDTLKGGPRAPPRNKMALYEQFSVFSHRY------  64

Qy   56   NSRSNNTSTLVHPGPS-SQPCGVERNLSVQHLDSSAANQ--------------ATEKFV  99
                    :|  |||    | :        :  |:                  ||||:|
Db   65   ------------SAAVPPAPSPAPPWGAQRPASA--VPSTSASQVGGGDRPIFPLFRVPSTEPVR  115

Qy   100  SQMSFMENVRSSAQ----------------------------------HDQRKMVREEEDF  126
          |: :: :|||:|:|
Db   116  S--SDQTNANSNGAAANGTIAESWRQRQSTHLKSKDTNAAGPPAEGNNSVGKKLANDDDF  173

Qy   127  AVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKP  186
                          ::|||  ::|||         |   |  ::     :
Db   174  TVPSVLYSGMP------PHSSQEKLTLFPTTSPCKSVPAKYSSTDK----------RRL  216

Qy   187  EVRD--QVKANARSGGFVISLDVSVTEEIDLE--KSASSHDRVNDYNASLRQESRNRLYR  242
          |||   :| ||||||||||||||||||||||||  |||||||||||||||:|||||||:
Db   217  EGMDASDVKSKGPSG---IKEKEPVQVRIDLEDKETTPSFQVLNDKTWSPDKLSSHM--  271

Qy   243  DGGKTRLKDTDNGAESHL----------ATENHSQEGHG---SPEDIDNDREYSKSRACASL  291
          ||| ||||:  :||                ||:||:|||      |:|:|
Db   272  ----DRLKKQHAEAESYQIRTRNENAVETQSPPKNGVSLLSKPYVDREQNGDSDLLGHGL  327

Qy   292  QQINE--------EASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAI  337
          ||:          |||||||||||||||||||||||||| ||||:||||||||
Db   328  RETGEKRKRSHHDVEQNDDLSDSS-VESLPGMEISPDDVVSAIGPKHFWKARRAI  381
```

```
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 253615
; LENGTH: 213
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_71038C.1.pep
US-10-424-599-253615

  Query Match             8.3%;  Score 306;  DB 12;  Length 213;
  Best Local Similarity   39.7%;  Pred. No. 3.5e-15;
  Matches   81;  Conservative  34;  Mismatches  69;  Indels  20;  Gaps  8;

Qy      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
          |:||:||||||||||||||||||||||||||||||||||||||||||:|
Db      1 MKRGNEEEKLGGPMFPRLHVNDAEKGGPRAPPRNKMALYEQFSIPSQKFNPR--LLPRKP 58

Qy     61 NTSTLVHPGPSSQPCGVERN-LSVQHLDSSAANQATEKFVSQMSFMENVRSS--AQHDQ 116
          |:|:|||   | | : : :  : ||:|||:||||||||||:||  ::  :: :
Db     59 NSSSNIVP-PPTQGNGHERSYVYPVRFSSQTPTHRAESYISRQS-DDGSRSNTSLVQLER 116

Qy    117 RKMVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQ 176
          |||||||:  ::    :||||: : : :     ::|||:|: ||||||||||:
Db    117 RKKVDDD-----IHVYTCSGIDQSNDKTMESVDGKKFTPFGARNFCYSV--AVQNDGDKD 169

Qy    177 NVCLATCSKP-----EVRDQVKAN 195
          |  | ||:|      |||||||||
Db    170 PTQFSSCSLPVDLRKDVRNGNEAN 193

RESULT 19
US-10-424-599-280208
; Sequence 280208, Application US/10424599
; Publication No. US2004031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 280208
; LENGTH: 203
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_95050C.1.pep
US-10-424-599-280208

  Query Match             8.0%;  Score 295;  DB 12;  Length 203;
  Best Local Similarity   38.9%;  Pred. No. 2.3e-14;
  Matches   82;  Conservative  26;  Mismatches  79;  Indels  24;  Gaps  5;

Qy      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
          |:||:||||||||||||||||||||||||||||||||||||||||||:|:
Db      1 MKRGNEEEKSAGPMFPRLHVNDABKGGPRAPPRNKMALYEQFSIPSQRFNPH--LPRKPN 58

Qy     61 NTSTLVHPGPSS-QPCGVERN-LSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRK 118
          | : |:||| |      : :  : :||||:||
Db     59 SSSNIVPPPTSSTEGNGRERSYISRQSDDGARSN----------------TSLLVQLERRK 102

Qy    119 MVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV 178
          ||||||||:|||||||||| : : :     :: :||: ||||||||||:
Db    103 KV----DDGIHVYTCSRIDQSNDKMLESFNGKKLTPFGARNFCYSVAVQNGGDKDTTQFG 158
```

```
Qy    179 CLATCSKPEVRDQVKANARSGGFVISLDVSV 209
          ::::  ||||||||| :: |:||::
Db    159 FLPADMRKDARKGNEANTHVSSSRQKLKLSV 189

RESULT 20
US-09-746-801A-24
; Sequence 24, Application US/09746801A
; Patent No. US2002008349441
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 179
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-746-801A-24

  Query Match             7.7%;  Score 281.5;  DB 9;  Length 179;
  Best Local Similarity   41.1%;  Pred. No. 2.1e-13;
  Matches   79;  Conservative  25;  Mismatches  61;  Indels  27;  Gaps  9;

Qy      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
          |||:||||||||||||||:||:: :|||||||||||||||||||||:||
Db      1 MKRGTGEKVMGPMFPRLNVNDTEKGGPRAPPRNKMALYEQLSIPSQRY-NPGDLPHNSS 59

Qy     61 NTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSF----MENVRSSAQHDQ 116
          | :: || |  ::  :  :|::||:||::|||||:||||||||         |||||:
Db     60 NSANLVLPHP-SQENEHERGVLFSR-QLPALRHPVEKPYGRSSGSNTPLREVKSKRQ--- 114

Qy    117 RKMVREEDFAVPVYINSR----RSQSHGRTKSGIEKE----KHTPMVAPSSHHSIRFQE 168
          :|| ||||:: ::         : :: ::|::|: | :: :
Db    115 ----TEKEDFRVPTFDNSKERAVNTEDYSKGTSDIDKRDSTLKRTDQL--SHVTPRENL 167

Qy    169 VNQTGS--KQNV 178
          |: ||   :||:
Db    168 VNTFGESHKTNI 179

RESULT 21
US-10-424-599-285624
; Sequence 285624, Application US/10424599
; Publication No. US2004031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With Plant Improvement
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 285624
; LENGTH: 110
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3847_99948C.1.pep
US-10-424-599-285624

  Query Match             7.0%;  Score 256.5;  DB 12;  Length 110;
  Best Local Similarity   65.0%;  Pred. No. 8.4e-12;
  Matches   52;  Conservative   7;  Mismatches  20;  Indels   1;  Gaps  1;

Qy      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
```

```
Db

Qy   1 MKRGKDDEKVMGPMFPRLHVNDTEKGPRAPPRNKMALYEQFSIPSQRF-NSGVLPLNPN 59
        ||||||||||:: |||||| :|||||||   :|||||||||:|||| :|: ||||||
Db

Qy  61 NTSTLVHPGPSSQPCGVERN 80
         :|||
Db  60 ISSNTVPPASSSLRTVPERN 79
```

RESULT 22
US-10-424-599-247729
; Sequence 247729, Application US/10424599
; Publication No. US20040031072A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa Thomas J
; APPLICANT: Kovalic David K
; APPLICANT: Zhou Yihua
; APPLICANT: Cao Yongwei
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53223)B
; CURRENT APPLICATION NUMBER: US/10/424,599
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 285684
; SEQ ID NO 247729
; LENGTH: 89
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: unsure
; LOCATION: (1)..(89)
; OTHER INFORMATION: unsure at all Xaa locations
; FEATURE:
; OTHER INFORMATION: Clone ID: PAT_MRT3B47_65729C.1.pep
US-10-424-599-247729

Query Match   7.0%; Score 255.5; DB 12; Length 89;
Best Local Similarity   66.2%; Pred. No. 7.5e-12;
Matches 53; Conservative 12; Mismatches 14; Indels 1; Gaps 1;

```
Qy 308 VDSISSIDVSPDDV-VGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPD 366
         : :||| |||||| ||:|||||  :||||||||||:||||||||||||:||||||||
Db   1 VTNLSSLIVSPDDVXVGILGQKHLWKARRKIANQQSVFAVQVFELHRLIKVQQLIAASPD 60

Qy 367 LLLDEISFLGKVSAKSYPVK 386
         ||: :|:||| |||:||||
Db  61 VLFEDGAFMGKYSLMGSPPK 80
```

RESULT 23
US-09-746-801A-25
; Sequence 25, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 49
; TYPE: PRT
; ORGANISM: Lycopersicon esculentum
US-09-746-801A-25

Query Match   6.3%; Score 231; DB 9; Length 49;
Best Local Similarity   85.7%; Pred. No. 2.4e-10;
Matches 42; Conservative   4; Mismatches   3; Indels 0; Gaps 0;

```
Qy   1 MKRGKDEEKILEPMFPPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRF 49
         ||||||:: |:|||| ||||||::||:|||||||||||||||:||||||
Db   1 MKRGKGEEKLMGPMFPRLHVNDTEKGGPKAPPRNKMALYEQLSIPSQRF 49
```

RESULT 24
US-10-425-114-43416
; Sequence 43416, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 43416
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: 700237614_FLI.pep
US-10-425-114-43416

Query Match   5.8%; Score 212; DB 12; Length 208;
Best Local Similarity   25.5%; Pred. No. 5.2e-08;
Matches 76; Conservative 34; Mismatches 72; Indels 116; Gaps 11;

```
Qy 411 EKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPQPPPSGNHQQ 470
        |||::.::: .  :|||:.:                QKSWCIP-APPS-----Q 53
Db   9 EKTEANQPSPSQDE--QAATNGDVAASMTPSDNK------QKSWCIP-APPS-----Q 53

Qy 471 WLIPVMSPSEGLIYKPHPGMAHTGHYGGIYGHYMPTPMVMPQYHPGMGFPPPNGYFPPY 530
        |||||||||||.:||||            :|||||
Db  54 WLIPVMSPSEGLVYKP-----YTGH-----------------------CPPVGSLLAPPF 85

Qy 531 GMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSY 590
Db  86 ------                                        ------FASY 89

Qy 591 PRARKSRQGST-GSSPSGPQGISGSKSFRPF----AAVDEDSNI------NNAPEQ 635
        ||:|       |||:||||:|.|::|||
Db  90 PTSSSTAGGDFMSSACGARLMSAPVYFPSFSMPAVSGSAVEQVSHVAASQHKRNSCSEA 149

Qy 636 TMTTTTTR--TTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYD 691
        |:|||  ||:| :|||:.|||||||||||:|..|.|:|||||||.|||.|||:|
Db 150 VLASRDSEVQGSSASSPASSETAAQPRVIRVVPHTARTASESAARIFRSIQMERKQND 207
```

RESULT 25
US-09-746-801A-19
; Sequence 19, Application US/09746801A
; Patent No. US20020083494A1
; GENERAL INFORMATION:
; APPLICANT: Wagner, et al.
; TITLE OF INVENTION: GENES REGULATING CIRCADIAN CLOCK FUNCTION AND PHOTOPERIODISM
; FILE REFERENCE: 1505-54357
; CURRENT APPLICATION NUMBER: US/09/746,801A
; CURRENT FILING DATE: 2000-12-20
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 19
; LENGTH: 185
; TYPE: PRT
; ORGANISM: Glycine max
US-09-746-801A-19

Query Match   5.5%; Score 202.5; DB 9; Length 185;
Best Local Similarity   33.3%; Pred. No. 2.4e-07;
Matches 69; Conservative 20; Mismatches 69; Indels 49; Gaps 9;

```
Qy   503 YMPTPMVMPQYHPGMG-----FPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQF 558
             :| | | | :    | | | | : | | | :|   |   :|        | | :| | |
Db    10 FMNPAYQFPPASHPVVGVSPFVPPASHTYFAPFGM--PVMN------QATSGSAVEQVNQF 61

Qy   559 GHPGNL-QN--------TQQQQRSDNEPAPQQQQPTKSYPPRARKSR-QGSTGSSP 605
            :|:|   |                  :|   | |  | | | | |   | | | |
Db    62 AAQGSHGQNGHSSVEGADFNTHHNQSSSNLPVQKNGARLHVKKSQALKERGLQGSTRSSP 121

Qy   606 SG-PQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVTQTTRDGGGVTRVIKV 664
            | | |       |  : |:::|
Db   122 SEMAQGIRARKI----------ADGSDAQSLSLHADETRQQ--------TQAIKV 158

Qy   665 VPHNAKLASENAARIFQSIQERRKRYD 691
         | | | | :|: | | || | | | | |::|
Db   159 VPHNRKSATESAARIVQSIQEERKQHD 185
```

Search completed: July 29, 2004, 08:32:20
Job time : 52 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 29, 2004, 08:20:25 ; Search time 14 Seconds
(without alignments)
2584.911 Million cell updates/sec

Title: US-10-719-885-2
Perfect score: 3669
Sequence: 1 MKRGKDEKILEPMFPRLHV........AARIFQSIQEERKRYDSSKP 695

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3669	100.0	695	1	ELF3_ARATH	O82804 arabidopsis
2	197	5.4	971	1	R3HD_HUMAN	Q15032 homo sapien
3	193.5	5.3	802	1	NAB3_YEAST	P38996 saccharomyc
4	189.5	5.2	738	1	YKF4_YEAST	P35732 saccharomyc
5	184	5.0	1048	1	SRA4_RAT	Q63627 rattus norv
6	179	4.9	971	1	CLA4_CANAL	O14427 candida alb
7	171.5	4.7	1596	1	MAM_DROME	P21519 drosophila
8	171	4.7	1157	1	SRA4_HUMAN	O95104 homo sapien
9	167	4.6	943	1	YM11_YEAST	P39523 saccharomyc
10	166.5	4.5	1429	1	EXPA_DROME	Q07436 drosophila
11	165.5	4.5	1983	1	TF20_MOUSE	Q9epg8 mus musculu
12	165	4.5	612	1	ADF1_CANAL	P46589 candida alb
13	164.5	4.5	629	1	PAB2_ARATH	P42731 arabidopsis
14	164.5	4.5	2716	1	OSA_DROME	Q8in94 drosophila
15	163.5	4.5	1516	1	NCO2_XENLA	Q9w705 xenopus lae
16	162.5	4.4	656	1	CPH1_CANAL	P43079 candida alb
17	162	4.4	816	1	ATX1_HUMAN	P54253 homo sapien
18	162	4.4	3726	1	ABF1_MOUSE	Q61329 mus musculu
19	161	4.4	1960	1	TF20_HUMAN	Q9ugu0 homo sapien
20	158	4.3	1089	1	Y553_HUMAN	Q9ukj3 homo sapien
21	158	4.3	2056	1	CBP1_CAEEL	P34545 caenorhabdi
22	156	4.3	1403	1	PROS_DROME	P29617 drosophila
23	155.5	4.2	964	1	YQKA_SCHPO	O74522 schizosacch
24	154.5	4.2	1902	1	SMF1_HUMAN	O14497 homo sapien
25	153	4.2	1693	1	SAS_DROME	Q04164 drosophila
26	152.5	4.2	734	1	YHU9_YEAST	P32900 saccharomyc
27	152	4.1	418	1	SSXT_MOUSE	Q62280 mus musculu
28	152	4.1	668	1	PAB5_ARATH	Q05196 arabidopsis
29	152	4.1	1586	1	SN22_HUMAN	P51531 homo sapien
30	151.5	4.1	2195	1	SC16_YEAST	P48415 saccharomyc
31	151	4.1	1082	1	NCO3_RAT	Q9epu2 rattus norv
32	150.5	4.1	727	1	BRC1_DROME	Q01295 drosophila
33	150.5	4.1	2441	1	CBP_MOUSE	P45481 mus musculu
34	150	4.1	1398	1	NCO3_MOUSE	O09000 m nuclear r
35	149.5	4.1	648	1	GLT0_WHEAT	P10387 triticum ae
36	149.5	4.1	1905	1	TAGB_DICDI	P54683 dictyosteli
37	149	4.1	653	1	SF01_MOUSE	Q64213 mus musculu
38	149	4.1	901	1	Y298_HUMAN	O15016 homo sapien
39	148.5	4.0	492	1	YNM4_YEAST	P53919 saccharomyc
40	148.5	4.0	1365	1	SU22_DROME	P25172 drosophila
41	148	4.0	467	1	YP25_CAEEL	Q09436 caenorhabdi
42	148	4.0	1128	1	ASP2_HUMAN	Q13625 homo sapien
43	147	4.0	707	1	FUB2_HUMAN	Q92945 homo sapien
44	147	4.0	840	1	YAO2_HUMAN	Q9y2k5 homo sapien
45	147	4.0	1157	1	BBC1_YEAST	P47068 saccharomyc
46	146.5	4.0	639	1	SF01_HUMAN	Q15637 homo sapien
47	146.5	4.0	1186	1	PKCB_HUMAN	Q9ulu4 homo sapien
48	146.5	4.0	2004	1	MYS3_HUMAN	Q92794 homo sapien
49	146	4.0	660	1	GLT3_WHEAT	P08488 triticum ae
50	146	4.0	1253	1	SHK2_HUMAN	Q9upx8 homo sapien
51	146	4.0	1474	1	SHK2_RAT	Q9qx74 rattus norv
52	145	4.0	873	1	RX_DROME	Q9w2q1 drosophila
53	145	4.0	1060	1	YN18_YEAST	P53836 saccharomyc
54	145	4.0	1230	1	ST20_CANAL	Q92212 candida alb
55	144.5	4.0	701	1	CG1_HUMAN	Q13495 homo sapien
56	144.5	3.9	1177	1	SP97_DICDI	Q95zg3 dictyosteli
57	144	3.9	418	1	SSXT_HUMAN	Q15532 homo sapien
58	143	3.9	648	1	KAPC_DICDI	P34099 dictyosteli
59	143	3.9	5262	1	MLL2_HUMAN	O14686 homo sapien
60	142.5	3.9	2063	1	NCO6_HUMAN	Q14686 h nuclear r
61	142	3.9	467	1	CBPA_DICDI	P35085 dictyosteli
62	142	3.9	861	1	PQ58_CAEEL	P34552 caenorhabdi
63	142	3.9	872	1	SCD5_YEAST	Q10172 schizosacch
64	142	3.9	1794	1	YDC9_SCHPO	Q9jks6 rattus norv
65	142	3.9	5085	1	RNC_HUMAN	Q9nrr4 homo sapien
66	141.5	3.9	1374	1	BRM_DROME	P25439 drosophila
67	141.5	3.9	1638	1	NCR1_HUMAN	O75376 homo sapien
68	141.5	3.9	2440	1	Z207_HUMAN	O43670 homo sapien
69	141	3.8	478	1	AXN2_HUMAN	Q9y2t1 homo sapien
70	140.5	3.8	843	1	CLMN_MOUSE	Q8c5w0 mus musculu
71	140.5	3.8	1052	1	CBP_HUMAN	Q92793 homo sapien
72	140.5	3.8	2442	1	HRX_HUMAN	Q03164 homo sapien
73	140.5	3.8	3969	1	VASP_MOUSE	P70460 mus musculu
74	140	3.8	375	1	BIB_DROME	P23645 arabidopsis
75	140	3.8	700	1	ABI3_ARATH	Q01593 rattus norv
76	140	3.8	720	1	FUB2_RAT	Q9cds8 mycobacteri
77	140	3.8	721	1	FTSH_MYCLE	P51521 drosophila
78	140	3.8	787	1	OVO_DROME	Q9bw71 homo sapien
79	140	3.8	1028	1	HIR3_HUMAN	Q9y7g2 candida alb
80	139.5	3.8	556	1	MIG1_CANAL	P10388 triticum ae
81	139.5	3.8	574	1	GLT5_WHEAT	Q15911 homo sapien
82	139.5	3.8	839	1	ABF1_HUMAN	Q9uqr1 homo sapien
83	139	3.8	3703	1	Z148_HUMAN	P30415 mus musculu
84	139	3.8	794	1	NKCR_MOUSE	P16753 human cytom
85	138.5	3.8	1453	1	VP40_HCMVA	Q96d71 homo sapien
86	138.5	3.8	708	1	REP1_HUMAN	Q92576 homo sapien
87	138.5	3.8	744	1	PHF3_HUMAN	Q8nez4 homo sapien
88	138	3.8	2039	1	MLL3_HUMAN	P36041 saccharomyc
89	138	3.8	4911	1	YKU4_YEAST	Q8r1s4 mus musculu
90	138	3.8	632	1	MTS1_MOUSE	Q62806 rattus norv
91	137	3.8	759	1	Z148_RAT	O60885 homo sapien
92	136.5	3.7	794	1	BRD4_HUMAN	Q9j104 mus musculu
93	136.5	3.7	1362	1	PAX3_MOUSE	P24610 triticum ae
94	136.5	3.7	1567	1	GLT4_WHEAT	P08489 triticum ae
95	136	3.7	479	1	YKZ6_CAEEL	P34333 caenorhabdi
96	136	3.7	838	1	GDA2_WHEAT	P04722 triticum ae
97	135.5	3.7	1780	1	CCT_DROME	O96433 drosophila
98	135.5	3.7	291	1	PMP3_MOUSE	P05143 mus musculu

ALIGNMENTS

RESULT 1

```
ELF3_ARATH
ID   ELF3_ARATH     STANDARD;      PRT;    695 AA.
AC   O82804; O04419; Q8L7A2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   EARLY FLOWERING 3 protein (Nematode responsive protein).
GN   ELF3 OR PYK20 OR AT2G25930 OR F17H15.25 OR T9L18.26.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   STRAIN=cv. C24; TISSUE=Shoot;
RX   MEDLINE=20035823; PubMed=10571046;
RA   Puzio P.S., Lausen J., Almeida-Engler J., Cai D., Gheysen G.,
RA   Grundler F.M.W.;
RT   "Isolation of a gene from Arabidopsis thaliana related to nematode
RT   feeding structures.";
RL   Gene 239:163-172(1999).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1), AND MUTANT ELF3-7.
RC   STRAIN=cv. Columbia, and cv. Wassilewskija;
RX   MEDLINE=21295524; PubMed=11402160;
RA   Hicks K.A., Albertson T.M., Wagner D.R.;
RT   "EARLY FLOWERING3 encodes a novel protein that regulates circadian
RT   clock function and flowering in Arabidopsis.";
RL   Plant Cell 13:1281-1292(2001).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Columbia;
RX   MEDLINE=20083487; PubMed=10617197;
RA   Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,
RA   Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblyum T.V.,
RA   Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,
RA   Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayam L.,
RA   Tallon L.J., Gill J.E., Adams M.D., Carrera A.J., Creasy T.H.,
RA   Goodman H.M., Somerville C.R., Copenhaver G.P., Preuss D.,
RA   Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,
RA   Venter J.C.;
RT   "Sequence and analysis of chromosome 2 of the plant Arabidopsis
RT   thaliana.";
RL   Nature 402:761-768(1999).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   STRAIN=cv. Columbia;
RX   MEDLINE=22954850; PubMed=14593172;
RA   Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J.,
RA   Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,
RA   Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,
RA   Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,
RA   Chan M.M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Ansari Y.,
RA   Arakawa T., Banh J., Banno F., Bowser L., Brooks S.Y., Carninci P.,
RA   Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,
RA   Hayashizaki Y., Johnson-Hopson C., Hsuan V.W., Iida K., Karnes M.,
RA   Khan S., Koesema E., Ishida J., Jiang P.X., Jones T., Kawai J.,
RA   Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,
RA   Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Empirical analysis of transcriptional activity in the Arabidopsis
RT   genome.";
RL   Science 302:842-846(2003).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=21295525; PubMed=11402161;
RA   Liu X.L., Covington M.F., Fankhauser C., Chory J., Wagner D.R.;
RT   "ELF3 encodes a circadian clock-regulated nuclear protein that
RT   functions in an Arabidopsis PHYB signal transduction pathway.";
RL   Plant Cell 13:1293-1304(2001).
CC   -!- FUNCTION: May be a transcription factor part of a circadian clock
CC       input pathway. Acts within a 'zeitnehmer' feedback loop and is
CC       involved in its own circadian regulation. Has no role in
CC       regulating circadian clock function in the dark. The activity of
CC       the protein may be decreased in long day conditions due to its
CC       interaction with phytochrome B (phyB). Can regulate the initiation
CC       of flowering independently of phyB. Also involved in responses to
CC       nematode parasitism, like the formation of the nematode feeding
CC       structure.
CC   -!- SUBUNIT: Interacts specifically with both Pr and Pfr forms of
CC       phytochrome B. May form a homodimer.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=O82804-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O82804-2; Sequence=VSP_004042, VSP_004043;
CC         Note=No experimental confirmation available;
CC   -!- INDUCTION: Expressed with a circadian rhythm showing a peak 14 to
CC       16 hours after sunrise regardless of daylength. Induced in roots
CC       after infection by nematodes. Upregulated by auxin and cytokinin
CC       and downregulated by abscisic acid and temperature stress.
CC   -!- CAUTION: Ref.1 sequence differs from that shown due to frameshifts
CC       in positions 437, 472 and 485.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; Y11994; CAA72719.1; ALT_FRAME.
DR   EMBL; AC004747; AAC31242.1; -.
DR   EMBL; AC005395; AAM15042.1; -.
DR   EMBL; AY136385; AAM97051.1; -.
DR   EMBL; BT000185; AAN15504.1; -.
DR   PIR; T02630; T02630.
DR   TRANSFAC; T05380; -.
KW   Nuclear protein; Transcription regulation;
KW   Phytochrome signaling pathway; Alternative splicing.
FT   DOMAIN       544    585       GLN-RICH.
FT   DOMAIN       442    539       PRO-RICH.
FT   DOMAIN       636    652       THR-RICH.
FT   VARSPLIC     339    339       N -> K (in isoform 2).
FT                                 /FTId=VSP_004042.
FT   VARSPLIC     340    695       Missing (in isoform 2).
FT                                 /FTId=VSP_004043.
FT   VARIANT      544    544       Q -> QQQQQQQQQQ (IN CV. WASSILEWSKIJA).
FT   MUTAGEN       66     73       MISSING: IN ELF3-7; CAUSES EARLY
FT                                 FLOWERING AND LONG HYPOCOTYL PHENOTYPES.
FT   CONFLICT      55     55       M -> R (IN REF. 1).
FT   CONFLICT     196    196       A -> R (IN REF. 1).
FT   CONFLICT     618    618       R -> G (IN REF. 1).
FT   CONFLICT     670    670       K -> KVVPHNAK (IN REF. 1).
SQ   SEQUENCE   695 AA;  77206 MW;  607A0720ED381C08 CRC64;

Query Match             100.0%;  Score 3669;  DB 1;  Length 695;
Best Local Similarity   100.0%;  Pred. No.7.6e-188;
Matches  695;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MKRGKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSN 60

Qy     61 NTSTLVHPGFSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 NTSTLVHPGFSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMV 120

Qy    121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 REEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
```

```
DR    InterPro; IPR001374; R3H.
DR    Pfam; PF01424; R3H; 1.
DR    SMART; SM00393; R3H; 1.
FT    DOMAIN    132    182    R3H.
FT    DOMAIN    461    480    POLY-PRO.
FT    DOMAIN    694    698    POLY-PRO.
SQ    SEQUENCE    971 AA;    107673 MW;    D07684D36955108 CRC64;

Query Match           5.4%;    Score 197;    DB 1;    Length 971;
Best Local Similarity    19.4%;    Pred. No. 0.0023;
Matches 142;    Conservative 94;    Mismatches 245;    Indels 250;    Gaps    33;

Qy    77    VERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRR    136
            :  :  :  :  ::  :    ::      : ::  :      :  ::  :
Db    93    LSRDSSQEYTDSTG--------IDLHEFLVNTLKNNPRDRMLLKLEQE--ILDFIGNNE    142

Qy    137    SQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEV----NQTGSKQNVCLATCSKPEV    188
             : ::  ::  :  :    :  : :
Db    143    S-------PRKKFPPM--TSYHRMLLHRVAAYFGLDHNVDQSGKSVIVNKTSNTRI    189

Qy    189    RDQV--------KANARSGGFVISLDVSVTEIDLE---------KSASSHDRVNDYNAS    231
                        ::: :  :  :  :  :  :  :  :
Db    190    PDQKFNEHIKDDKGEDFQKRYILKRDNSSFDKDDNQVIRLKDDRRSKSIEERREEY--    246

Qy    232    LRQESRNRLYRDGG---------------KTRLKDTD-----NGAESHLATENHSQEG    269
                :    :    :          :      :
Db    247    -QRARDRIFSQDSLCSQENYIIDKRLQDEDASSTQQRRQIFRVNKDASGRSTNSHQ--    301

Qy    270    HGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSM--------VDSISSIDV-SPD    319
             ::  :  : :: :        :  :  ::  :
Db    302    ---SSTENELKYSEPRPWSSTD-----------------SDSSLRNLKPAVTKASSFSGISVLTRG    347

Qy    320    DVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLDEISFLGKVS    379
            ::  :                :                :      :
Db    348    DSSG--------SSKSIG--------RLSK-----TGQPFINPDGSPVVVNPP    379

Qy    380    HVVVKQRGDSEKTDQH    416
            ::
Db    380    MTQQPVRSQVPGP--QPPLPAPPGQPAANHIFSQQDNLGSQFSHMSLARQPSADGSDPH    437

Qy    417    KMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPP--    464
             :    : : :
Db    438    AAMFQSTVVL-----QSPQQSGYIMTAAPPPHPPP-----PPPPPPPPPLPPGQPVPT    486

Qy    465    ----SGNHQQWLIPVMSP----SEGLIYKPHGM---AHTGHYGGYYGHYMPTPMVMPQ    512
                :  ::    :  :  :  :      :
Db    487    AGYPAGGH--------PVSQPVLQQQGYIQQPSPQMPACYCAPGHYHSSQPQYRPVPSV--H    538

Qy    513    YHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ---QQPNEQMNQFGHPGNLQN-    566
                 :: :    : :    :  :      :  :  ::  ::
Db    539    YNSHLNQPLPQPAQQTGYQVIP----------NQQQNYQGIVGVQQPQSQSLVSGQPNSIGNQ    591

Qy    567    TQQQRSDNEPAPQQQQPTKSYPRARKSRQGTSGSPSGPQGISGS    614
             :    :      :  :
Db    592    IQGVIPYTSVPTYQVSLPQGSQGIPHQTYQQPVMFP----NQSNQGSMPT----TGM    641

Qy    615    KSFRPFAAVDEDSNINNA----------PEQTMTTTTTTT----RTTVTQTTRDGGG    657
             : :  :    :                 :              :::
Db    642    PVYYSVIPPGQQNLSSSVGYLQHPGSEQVQFPRTTSPCSSQQLQHQCTAGPPPPPGGG    701

Qy    658    VTRVIKVPHN    668
                ::
Db    702    MVMMQLSVPNN    712


RESULT 3
NAB3_YEAST
ID    NAB3_YEAST    STANDARD;    PRT;    802 AA.
AC    P38996;
DT    01-FEB-1995 (Rel. 31, Created)
DT    01-FEB-1995 (Rel. 31, Last sequence update)
DT    10-OCT-2003 (Rel. 42, Last annotation update)
DE    Nuclear polyadenylated RNA-binding protein NAB3.
```

```
Qy    181    ATCSKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSASSHDRVNDYNASLRQESRNRL    240
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181    ATCSKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSASSHDRVNDYNASLRQESRNRL    240

Qy    241    YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD    300
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241    YRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASD    300

Qy    301    DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL    360
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301    DVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKL    360

Qy    361    IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES    420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361    IAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMES    420

Qy    421    SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPQPPPSSGNHQQWLIPVMSPSE    480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421    SAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPQPPPSSGNHQQWLIPVMSPSE    480

Qy    481    GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPY    540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481    GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPY    540

Qy    541    CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS    600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541    CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGS    600

Qy    601    TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVTVTQTTRDGGGVTR    660
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601    TGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVTVTQTTRDGGGVTR    660

Qy    661    VIKVPHNAKLASENAARIFQSIQERKRYDSSKP    695
             ||||||||||||||||||||||||||||||||
Db    661    VIKVPHNAKLASENAARIFQSIQERKRYDSSKP    695


RESULT 2
R3HD_HUMAN
ID    R3HD_HUMAN    STANDARD;    PRT;    971 AA.
AC    Q15032;
DT    15-JUL-1998 (Rel. 36, Created)
DT    15-JUL-1998 (Rel. 36, Last sequence update)
DT    28-FEB-2003 (Rel. 41, Last annotation update)
DE    R3H domain protein 1.
GN    R3HDM OR KIAA0029.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=Bone marrow;
RX    MEDLINE=96051387; PubMed=7584026;
RA    Nomura N., Miyajima N., Sazuka T., Tanaka A., Kawarabayasi Y.,
RA    Sato S., Nagase T., Seki N., Ishikawa K.-I., Tabata S.;
RT    "Prediction of the coding sequences of unidentified human genes. I.
RT    The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by
RT    analysis of randomly sampled cDNA clones from human immature myeloid
RT    cell line KG-1.";
RL    DNA Res. 1:27-35(1994).
-!-  SIMILARITY: Contains 1 R3H domain.

CC    -------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -------------------------------------------------------------------
DR    EMBL; D21852; BAA04878.1; -.
DR    Genew; HGNC:9757; R3HDM.
```

```
Best Local Similarity   20.2%;   Pred. No. 0.0028;
Matches  128;  Conservative   73;  Mismatches 216;   Indels  217;   Gaps  29;

Qy      81  LSVQHLDSSANQATEKFVSQMSFMENVRSSAQHDQRKM---------VREEEDFAVPV  130
                :  |:  :  :.       .:   : :        |:|
Db     253  LSVPPHDSTISTTASASATSGARSNDQRKPPLSDAQRRMRFPRADLSKPITEEEHDRYAA  312

Qy     131  YINSRRSQSHGRTK-----------------------SGIEKEKTPMVAPSSHHSIRFQEVN  170
              |: |                                 .||:|
Db     313  YL----HGENKITEMHNIPPKSRLFIGNLPLKNVSKEDLFRIFSPYGH---------IM  358

Qy     171  QTGSKQNVCLATCSKPE-VRDQVKANARSGGFVISLDVSVTEEIDLEKSASS----------H  222
                  : ::      :        :::  ::: ::
Db     359  QINIKNAFGFIQFDNPQSVRDAIECE----------SQEMNFGKKLILEVSSSNARPQFDHG  410

Qy     223  DRVNDYNASLRQESRNRLYRDGGKTRLKDTDNG------------AESHLAT  262
              :  :  .:.:       |:|
Db     411  DHGTNSSSTFISSAKRPFQTESGD--MYNDDNGAGYKKSRRHTVSCNIFVKRTADRTYAI  468

Qy     263  E--NHSQEGHGSPEDINDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDD  320
              :     |:: ::        : :     :::::|
Db     469  EVPNRFRDGTG----LETDMIFLKPRMELG-KLINDAAYNGWNGVVLVNKTHNVDV---  519

Qy     321  VVGILGQKRFWRARKAIANQQRVFAVQLFELIRLIKVQKLIAASPDLLLDEISFLGKVSA  380
                                                     |:     :
Db     520  ---------QTFYKGSQGET----------------------KFDEYISISAD--DAVAIFNNIK-  551

Qy     381  KSYPVKKLLPSEFIVKPPLPHVVVKQRGDSEKTDQHKMESSAENVGRLSNQGHHQQSNY  440
                                                           |||| ::
Db     552  ------------------------------NNRNNSRPTDYRAMS---------HQQNIY  572

Qy     441  ----MPFANNPPASPAP------NGYCFPP------QPPPSGNHQQWLIPVMSPSEGLIYKPH  487
                   |  |: |   |                       |
Db     573  GAPPLPVPNGPAVGPPPQTNYYQGYSMPPPQQQQQPYGNYG---MPPPSHDQG--YGSQ  627

Qy     488  PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGF---------PPPGNGYFPP--------  529
              | |: :   :   ::    |:  :          |||
Db     628  PPIPMNQSYGRYQTSIPPP---PQQQIPOGYGRYQAGPPPQ---PPSQTPMDQQQLLS  680

Qy     530  --YGMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQOQPT  587
                  ::: ::||||||||          ||||:|
Db     681  AIQNLPPNVVSNLLSMAQQQQQQPHAQQQLV---GLIQSMQGQ---------APQQQQQL  729

Qy     588  KSYPRARKRQGSTGSSPSGPQGISGSKSFRPFA  621
                    :   ::  :::  |: ||:::|
Db     730  GGYSMNSSSPPPMSTNYNG-QNISAKPSAPPMS  762
```

RESULT 4
YKF4_YEAST
ID   YKF4_YEAST   STANDARD;   PRT;   738 AA.
AC   P35732;
DT   01-JUN-1994 (Rel. 29, Created)
DT   01-JUN-1994 (Rel. 29, Last sequence update)
DT   01-FEB-1996 (Rel. 33, Last annotation update)
DE   Hypothetical 84.0 kDa protein in NUP120-CSE4 intergenic region.
GN   YKL054C OR YKL308.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c;
RX   MEDLINE=94378723; PubMed=8091862;
RA   Rasmussen S.W.;
RT   "Sequence of a 28.6 kb region of yeast chromosome XI includes the
RT   FBA1 and TOA2 genes, an open reading frame (ORF) similar to a
RT   translationally controlled tumour protein, one ORF containing motifs
RT   also found in plant storage proteins and 13 ORFs with weak or no
RT   homology to known proteins.";
RL   Yeast 10:563-S68(1994).
CC   -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   used for commercial purposes. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).

GN   NAB3 OR YPL190C.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Wilson S.M., Oberdorf A.M., Datar K.V., Swedlow J.R., Paddy M.R.,
RA   Swanson M.S.;
RL   Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313271; PubMed=9169875;
RA   Bussey H., Storms R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA   Araujo R., Aparicio A., Barrel B.G., Badcock K., Benes V.,
RA   Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA   Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA   Dietrich F.S., Delius H., DiPaolo H., Dubois E., Duesterhoeft A.,
RA   Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA   Hall J., Hebling U., Heumann K., Hilbert H., Hillier L.W.,
RA   Hunicke-Smith S., Hyman R., Johnston M., Kalman S., Kleine K.,
RA   Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA   Marathe R., Messenguy F., Mewes H.-W., Mirtipati S., Moestl D.,
RA   Mueller-Auer S., Namath A., Nentwich U., Oefner P., Pearson D.,
RA   Petel F.X., Pohl T.M., Purnelle B., Schafer M., Scharfe M.,
RA   Scherens B., Schramm S., Schroeder M., Sdicu A.M., Tettelin H.,
RA   Urrestarazu L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA   Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA   Zhong W.W., Zollner A., Vo D.H., Hani J.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome XVI.";
RL   Nature 387:103-105(1997).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
DR   EMBL; U05314; AAA81910.1; -.
DR   EMBL; Z73546; CAA97903.1; -.
DR   PIR; S48529; S48529.
DR   GermOnline; 144172; -.
DR   SGD; S0006111; NAB3.
DR   GO; GO:0005654; C:nucleoplasm; IDA.
DR   GO; GO:0008143; F:poly(A) binding; IDA.
DR   GO; GO:0006357; P:regulation of transcription from Pol II pro. . .; IGI.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   RNA-binding; Nuclear protein.
FT   DOMAIN   30   101   RNA-BINDING (RRM).
FT   DOMAIN   38   71   POLY-GLU.
FT   DOMAIN   87   93   POLY-GLU.
FT   DOMAIN   101   106   POLY-ASP.
FT   DOMAIN   108   115   POLY-GLU.
FT   DOMAIN   116   127   POLY-ASP.
FT   DOMAIN   128   137   POLY-GLN.
FT   DOMAIN   603   608   POLY-GLN.
FT   DOMAIN   644   648   POLY-PRO.
FT   DOMAIN   698   703   POLY-GLN.
FT   DOMAIN   723   728   POLY-GLN.
FT   DOMAIN   765   768   POLY-PRO.
FT   DOMAIN   769   784   POLY-GLN.
SQ   SEQUENCE   802 AA; 90438 MW; CD6D2C7F24A44993 CRC64;

Query Match   5.3%; Score 193.5; DB 1; Length 802;

```
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC
DR   EMBL; X75781; CAA53418.1; -.
DR   EMBL; Z28054; CAA81890.1; -.
DR   PIR; S37876; S37876.
DR   GermOnline; 139810; -.
DR   SGD; S0001537; VID31.
DR   GO; GO:0005634; C:nucleus; IMP.
DR   GO; GO:0006974; P:response to DNA damage stimulus; IMP.
DR   GO; GO:0006511; P:ubiquitin-dependent protein catabolism; IDA.
KW   Hypothetical protein.
FT   DOMAIN      382    718       GLN-RICH.
FT   DOMAIN      574    580       POLY-ALA.
FT   DOMAIN      708    716       POLY-ALA.
SQ   SEQUENCE   738 AA;  83973 MW;  05734E2D0B7389AC CRC64;

Query Match             5.2%;  Score 189.5;  DB 1;  Length 738;
Best Local Similarity  17.5%;  Pred. No. 0.0041;
Matches  128;  Conservative   99;  Mismatches  248;  Indels  257;  Gaps   29;

Qy     19 HVN-DADKGGPRAPPRNKMALYEQLSIPSQRFGDFGT-MNSRSNNTSTLVHPGPSSQPCG 76
          ||:         | :||  |:||: |:         |   :      ::|| |  |
Db    142 HVNLDKGKHVPSKPVSNTTSWAAAVSVDTK----HDVPQDSNDNNNEELEAQGQQAQEKN 197

Qy     77 VERNLSVQHLDSSAANQATRKFVSQMSFMENVRSSAQHDQRKMVREEEDFAVPVVINSRR 136
          |:                                                  :|::|
Db    198 QEKSQEEQQQQEGHNNKEEHKQIEQPS------------------------LSSKK 229

Qy    137 SQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANA 196
          :|::|          |       |  :||                ||  || ||:|
Db    230 TTS--RTSASQPKKMSWAAIATPKPKAVKKTE----SPLENVAEL--KKEISDIKK---- 277

Qy    197 RSGGFVISLDLVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGA 256
          :| |                                        :|
Db    278 --------------DDQKSEASEEKVNEQETSAQEQEE------------------- 302

Qy    257 ESHLATENHSQEGHGS$PEDIDNDREYSKSRACASLQQINEEASDVSDDSMVDSISSIDV 316
          :|   : |||  | : |: | |   |:    || | |::  ||||  ::   |:
Db    303 --------------TAEPSEENEDRVPEVDGEEVQERAE---KKEQVKEEQTAEELEQEQDNVA---- 350

Qy    317 SPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLLIKVQKLIAASPDLLLDEISFLG 376
          ||:|:| :| :: ||
Db    351 -------------------------VEISAVISEPPE--DQANTVP 381

Qy    377 KVSAKS--------YPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSSAENVV 426
          ||: :         :|:|| | | | |:| |:|||
Db    382 QPQQQSQQPQQPQQPQQ--QQOPQQPQQPQQQQQQQQQQPVQAQAOQE 440

Qy    427 GRLS---------NQGHHQQSNYM---PFANNPPASPAPNGY----CFPPQPPPS 465
          |  :          ::|  ||:|       || | ||
Db    427 GRLS---------NQGHHQQSNYM---PFANNPPASPAPNGY----CFPPQPPPS 465

Qy    441 EQLSQNYYTQQQQQYAQQQHQLQQQYLSQQQYAQQQQQHPQPQSQQPQSQQSPQSQKQ 500
          |  ||||||  ||||  |||  |    || ||  |    | | | || |  || ||
Db    427 EQLSQNYYTQQQQQYAQQQHQLQQQYLSQQQYAQQQQQHPQPQSQQPQSQQSPQSQKQ 500

Qy    466 GNH---QQWLI-----------PVMSPSEGLIY----------------------------- 484
          ||: :|
Db    501 GNNVAAQQYMYQNQFPGYSYPGMFDSQYAYGGQYQYQQLAQNNAQTSGNANQYNFQQQYG 560

Qy    485 ------------KPHPGMAHT----GHYGG-----YYGHY--MPTPMVMPQQYHP 515
          ||
Db    561 QAGANTAAANLTSAAAAAASPATAHAQPQQQQYGGSFMPYYAHFYQQSFPYGQPQYGV 620

Qy    516 GMGFP--PPGNGYFPPYGMMPTIMNPYCSSQQQQPNEQMNQFGHPCNLQNTQQQQQR 573
          | | |    ||||  || |
Db    621 AGQYPYQLLPKNNY---NYYQTQNGQEQQQSPNQGVAQHSEDSQQKQSQQQQQQ 669

Qy    574 --SDNEPA-PQQQQQPTKSYPRARKSRQSTGSSGPGPGISGSKSFRPF 620
          |:     |      |     |  |  |  |  |
Db    670 QPQGQPQPEVQMQNGQPVNPQQQMQFQQYYQFQQQQQAAAAQGV-----PY 722
```

RESULT 5

```
SRA4_RAT
ID   SRA4_RAT                STANDARD;      PRT;  1048 AA.
AC   Q63627;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   CTD-binding SR-like protein RA4 (Fragment).
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Hippocampus;
RX   MEDLINE=96293459; PubMed=8692929;
RA   Yuryev A., Patturajan M., Litingtung Y., Joshi R.V., Gentile C.,
RA   Gebara M., Corden J.L.;
RT   "The C-terminal domain of the largest subunit of RNA polymerase II
RT   interacts with a novel set of serine/arginine-rich proteins.";
RL   Proc. Natl. Acad. Sci. U.S.A. 93:6975-6980(1996).
CC   -!- FUNCTION: May act to physically and functionally link
CC       transcription and pre-mRNA processing (By similarity).
CC   -!- SUBUNIT: Interacts with the repetitive C-terminal domain (CTD) of
CC       RNA polymerase II (By similarity).
CC   -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U49058; AAC52660.1; -.
DR   PIR; T31425; T31425.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   RNA-binding.
FT   NON_TER       1      1
FT   DOMAIN        3      6       POLY-PRO.
FT   DOMAIN      155    158       POLY-PRO.
FT   DOMAIN      422    496       RNA-BINDING (RRM).
FT   DOMAIN      630    638       POLY-PRO.
FT   DOMAIN      655    661       POLY-PRO.
FT   DOMAIN      834    857       POLY-GLN.
SQ   SEQUENCE   1048 AA;  113701 MW;  F5CE8FD309D6883E CRC64;

Query Match             5.0%;  Score 184;  DB 1;  Length 1048;
Best Local Similarity  26.8%;  Pred. No. 0.012;
Matches   84;  Conservative   25;  Mismatches   94;  Indels  110;  Gaps   16;

Qy    389 LPSEFL------------------VKPP----------LPHVVVKQRGDSEKTDQHKMESSAENVVGRL 429
          ||||                    |||               :::  |:|
Db    649 LPPGFLPPGPPPPTTPPVSIPPHTPPISIPNLVSGARGNAESADSAKMYGSA------- 701

Qy    430 SNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPP----------SGNHQQWLIPVMSPSEGL 482
          | |:       | | | |            | |  ||  |  || |  |  : | || | |
Db    702 ------------------------------GPPA--APTSLPTTPPVTQVSLLGTQGVAPGPVIGLQAPSTGLL 743

Qy    483 --------IYKPHGMAHTGHYGGYYGHYMPTPMVMPQ---------YHPGMGFPPPGPNGYF 527
          |  |  | | | || | | ||         ::| | |
Db    744 GGRPGLIPLQRP-PGMPPP------------HLQRFPMMPPRPMPPEHMMHRG---PPPGPGGFA 791
```

Qy   621 AAVDEDSNINNA 632

Qy   723 GYNGYDYNSKNS 734

```
Qy   528  --PPYGM-----------------MPTIMNP------YCSSQQQQQQPNEQMNQFGHPGNLQ  565
               :|||||                  |||  ::||| :|||| :|
Db   792  MPPPHGMKGPFPFPHGPFVRPGGMPGLGGPGPGGSEDRDGRQQQPQQQQQ------QQ     845

Qy   566  NTQQQQQRSDNEPAPQQ--QQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAA  622
               ||||:  :|:|||| |||||
Db   846  QQQQQQQQQQQPPPQQSCTQQQPAPSQQPAPAQQQPQQQFRNDNRQQFNSGRDQER-FGR  904

Qy   623  VDEDSNINNAPEQ  635
               : :||||
Db   905  RSFGSRVENDRER  917


RESULT 6
CLA4_CANAL
ID   CLA4_CANAL    STANDARD;     PRT;   971 AA.
AC   O14427;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Serine/threonine-protein kinase CLA4 (EC 2.7.1.-).
GN   CLA4.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=9411146; PubMed=9259554;
RA   Leberer E., Ziegelbauer K., Schmidt A., Harcus D., Dignard D., Ash J.,
RA   Johnson L., Thomas D.Y.;
RT   "Virulence and hyphal formation of Candida albicans require the
RT   ste20p-like protein kinase CaCla4p.";
RL   Curr. Biol. 7:539-546(1997).
CC   -!- FUNCTION: Essential for virulence and morphological switching
CC       (hyphal formation) of C.albicans.
CC   -!- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.
CC       STE20 subfamily.
CC   -!- SIMILARITY: Contains 1 CRIB domain.
CC   -!- SIMILARITY: Contains 1 PH domain.
CC   -------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   used by                                                    and for commercial
CC   modified and this statement is not removed. Usage by  and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------
DR   EMBL; U87996; AAB68613.1; -.
DR   HSSP; Q63450; 1A06.
DR   InterPro; IPR000095; PAKbox/Rhobndng.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000719; Prot_kinase.
DR   InterPro; IPR008271; Ser_thr_pkin_AS.
DR   InterPro; IPR002290; Ser_thr_pkinase.
DR   InterPro; IPR001245; Tyr_pkinase.
DR   Pfam; PF00786; PBD; 1.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00069; pkinase; 1.
DR   PRINTS; PR00109; TYRKINASE.
DR   ProDom; PD000001; Prot_kinase; 1.
DR   SMART; SM00285; PBD; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00220; S_TKc; 1.
DR   PROSITE; PS50108; CRIB; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR   PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR   PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW   Transferase; Serine/threonine-protein kinase; ATP-binding.
FT   DOMAIN     67    178  PH.
FT   DOMAIN    231    244  CRIB.
FT   DOMAIN    680    935  PROTEIN KINASE.
FT   DOMAIN     16     19  POLY-PRO.
FT   DOMAIN    210    213  POLY-SER.
FT   DOMAIN    311    315  POLY-ASN.
FT   DOMAIN    384    389  POLY-ASN.
FT   DOMAIN    392    395  POLY-ASN.
FT   DOMAIN    418    421  POLY-PRO.
FT   DOMAIN    453    462  POLY-GLN.
FT   DOMAIN    468    476  POLY-GLN.
FT   DOMAIN    572    576  POLY-GLN.
FT   NP_BIND   686    694  ATP (BY SIMILARITY).
FT   BINDING   710    710  ATP (BY SIMILARITY).
FT   ACT_SITE  803    803  BY SIMILARITY.
SQ   SEQUENCE   971 AA;  106889 MW;  AD6F0DBBC6CF624B CRC64;

Query Match     4.9%;  Score 179;  DB 1;  Length 971;
Best Local Similarity   25.0%;  Pred. No. 0.02;
Matches   93;  Conservative   37;  Mismatches  132;  Indels  110;  Gaps   19;

Qy   377  KVSAKSYPVKKLLPSEFLVKPPLPHVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQ  436
               | |||   ||  |: |:  ::|| ::|||   : :::
Db   337  KPPAKS-TVSQFKPSRAAPKPPTPYHLTQLNGSS-----HQHTSSS----GSLPSSGNNN  386

Qy   437  QSNYMPPFANNPPASPAPN----GYCFP-----PQPPSG--------NH------QQWL  472
               |   :|||      :       ||   ||
Db   387  NNNSTNNNNTKNVSPLNLMNKSELIPARRAPPPTSTSSDTYSNKNHQDRSGYEQORQ    446

Qy   473  IPVMSPSEGLIYKPH------------------PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFP  520
               ::|   |      :|                 |    |  |||::: |
Db   447  QRTDSSQQQQQQKQHQYQQKSQQQQQPQQPLSSHQGG-TSH--IPKQVPPTLPSSG-P     501

Qy   521  P--PGNGYFPYGMMPTI-----MNPYCSSQQQQQQQPNEQMNQFGHPGNCNLQNTQQQQQR  573
               |  |  || || || |       ::| : :::   : ::::|||
Db   502  PTQAASGKSMPSKIHPDLKIQQGTNNYIKSSGTDANQVDGDAKQFIKPNLQSKKSQQQL    561

Qy   574  SDNEPAP----QQQQPTKSYPRARKSRQGSTGSS---------------PSGPQGISGSKSF  617
               : :||:  ||||           |||:
Db   562  ASKQPSPPSSQQQQKPMTSH-----------GLMGTSHSVTKPLNPVNDPIKPLNLKSSKSK  613

Qy   618  RPFAAVDEDSNINNAPEQTMTTTTTTTTRTTVTQTTRDGGGVTRVIKVVPHNAKLASENAA  677
               :::||
Db   614  E---ALNETSGVSKTPSPTDKSNKPTAPAS-----------GPAVTKTAK-----        649

Qy   678  RIFQSIQEERKR  689
               ::|||:|
Db   650  ----QLKKRER  657


RESULT 7
MAM_DROME
ID   MAM_DROME    STANDARD;     PRT;   1596 AA.
AC   P21519;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Neurogenic protein mastermind.
GN   MAM.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Canton-S;
RX   MEDLINE=91065516; PubMed=1701150;
RA   Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA   Yedvobnick B.;
RT   "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT   unusually rich in amino acid homopolymers.";
RL   Genes Dev. 4:1688-1700(1990).
CC   -!- FUNCTION: May have a regulatory function possibly in association
```

```
Qy    331  WRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLDE--------------------------  371
                ||||  :|
Db    486  -----------------------------------FPGPPDLLGDDNSENNDTFKDLINNLQDFN  515

Qy    372  ISFLGKVSAKSYPVKKLLPSEFL-VKPP------------------------------LPHVV  403
                ||||| :|:|    |  | : || |
Db    516  PSFLDGFDEK-PLLDIKTEDGIKVEPPNAQDLINSLNVKSEGLGHGFGGFGLGLDNPG  573

Qy    404  VKQRGDSE-----KTDQHKMESSAENVVGRLSNQGH---------------------------HQQSN  439
                |:|:        :|||
Db    574  MKMRGGNPGNQGGFPNGPNGGTGGAPNAGGNGGNSGNLMSEHPLAAQTIKQMAEQHQHKN  633

Qy    440  YMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGY  499
                :     :    |      :|     ||||||||||||||||||||||||||||
Db    634  AM------------GGMGGFPRPHGMNPQQQQQQQQQQQQQQHGQMMGGQGQPGRY  681

Qy    500  ----------------------YGHYMPTPM------------VMPQYHPGMGFPPPGNGYFPPYGMMP  534
                                                    :|:||||||||||||||||||||||
Db    682  NDYGGGFPNDFGLGPNGPQQQQAQQQQPQQQHLPPQFHQQKG-PGPGAGMNVQQQNFLD  740

Qy    535  TIMNPYCSSQ--------------QQQQQQPNEQMNQ---FGHPG-NLQNTQQ----QQQR  573
                :|:                 ||||||           :|:::           |||
Db    741  IKQELFYSSQNDFDLKRLQQQAMQQQQQQHHQQQPKMGGVPNFNKQQQQQQVPQQQL  800

Qy    574  SDNEPAPQQQQQPTKSY-----------PRARKS------RQGSTGSSP----------SGP  608
                                         ||||||        :||||||||
Db    801  QQQQQQQQQQQQQQYSPFSNQNPNAAANFLNCPRGGPNGNQQPGNLAQQQQOQQPGAGP  860

Qy    609  QGISGSKSFRPFAAVDEDSNIN-------NAPEQTMTTTTTTTRTTVTQT-----TRDG  655
                |||:::|||  ::|           :|:|||::|||||||: |||||||
Db    861  Q------QQQRGNAGNGQQNNPNTGPGNTPNAPQQQQQQQSTTTTLQMKQTQQLHISQQG  916

Qy    656  GGVTRV-IKVVPH---------NAKLASENAARIFQSIQEERKRYDSSKP  695
                |     :|:::          |::  |          :|
Db    917  GGAQGIQVSAGQHLHLSGDMKSNVSVAAQQG--VFFSQQQAQQQQQQQQP  964
```

```
RESULT 8
SRA4_HUMAN
ID    SRA4_HUMAN              STANDARD;       PRT;   1157 AA.
AC    O95104; Q9UFM1; Q9ULP8;
DT    30-MAY-2000 (Rel. 39, Created)
DT    16-OCT-2001 (Rel. 40, Last sequence update)
DT    10-OCT-2003 (Rel. 42, Last annotation update)
DE    CTD-binding SR-like protein RA4 (Fragment).
GN    KIAA1172.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Tassone F., Villard L., Gardiner K.;
RT    "Sequence, genomic organization and map localization of the human SR
RT    protein gene rA4.";
RL    Submitted (AUG-1997) to the EMBL/GenBank/DDBJ databases.
RN    [2]
RP    SEQUENCE FROM N.A.
RX    MEDLINE=20289799; PubMed=10830953;
RA    Hattori M., Fujiyama A., Taylor T.D., Watanabe H., Yada T.,
RA    Park H.-S., Toyoda A., Ishii K., Totoki Y., Choi D.-K., Groner Y.,
RA    Soeda E., Ohki M., Takagi T., Sakaki Y., Taudien S., Blechschmidt K.,
RA    Polley A., Menzel U., Delabar J., Kumpf K., Lehmann R., Patterson D.,
RA    Reichwald K., Rump A., Schillhabel M., Schudy A., Zimmermann W.,
RA    Rosenthal A., Kudoh J., Shibuya K., Kawasaki K., Asakawa S.,
RA    Shintani A., Sasaki T., Nagamine K., Mitsuyama S., Antonarakis S.E.,
RA    Minoshima S., Shimizu N., Nordsiek G., Hornischer K., Brandt P.,
RA    Scharfe M., Schoen O., Desario A., Reichelt J., Kauer G., Bloecker H.,
RA    Ramser J., Beck A., Klages S., Hennig S., Riesselmann L., Dagand E.,
RA    Wehrmeyer S., Borzym K., Gardiner K., Yaspo M.-L.;
RL    Lehrach H., Reinhardt R., Yaspo M.-L.;
RT    "The DNA sequence of human chromosome 21.";
RL    Nature 405:311-319(2000).
```

```
CC      with the N gene product.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DEVELOPMENTAL STAGE: During early neurogenesis mam products are
CC       ubiquitously located. During later stages they accumulate in the
CC       central nervous system.
CC   -!- MISCELLANEOUS: The protein has many AA homopolymeric domains: 21
CC       poly-Gln runs (from 5 to 16 AA in length) 4 poly-Gly (6 to 10
CC       AA), 3 poly-Asn (3 x 5 AA), 1 poly-Ala (10 AA) and 1 poly-Thr (5
CC       AA) runs.
CC   -!- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC       YEAST AND MAMMALIAN REGULATORY PROTEINS.
CC
CC   ----------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics  and the  EMBL outstation -
CC       the European Bioinformatics Institute.  There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed.  Usage  by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR    EMBL; X54251; CAA38152.1; -.
DR    PIR; A36391; A33106.
DR    FlyBase; FBgn0002643; mam.
DR    GO; GO:0005634; C:nucleus; IDA.
DR    GO; GO:0007500; P:mesoderm cell fate determination; IMP.
DR    GO; GO:0007399; P:neurogenesis; NAS.
KW    Neurogenesis; Nuclear protein; Repeat.
FT    DOMAIN       20     84       GLN-RICH.
FT    DOMAIN      127    190       ARG/LYS-RICH (BASIC).
FT    DOMAIN      196    219       GLN-RICH.
FT    DOMAIN      259    304       ASN-RICH.
FT    DOMAIN      355    388       GLY/ASN-RICH.
FT    DOMAIN      392    406       GLN-RICH.
FT    DOMAIN      407    440       GLY-RICH.
FT    DOMAIN      651    671       GLN-RICH.
FT    DOMAIN      700    714       GLN-RICH.
FT    DOMAIN      759    816       GLN-RICH.
FT    DOMAIN      987    996       5 X 2 AA TANDEM REPEATS OF G-V.
FT    DOMAIN     1060   1079       ALA-RICH.
FT    DOMAIN     1092   1107       8 X 2 AA TANDEM REPEATS OF V-G.
FT    DOMAIN     1237   1252       7 X 2 AA TANDEM REPEATS OF G-V.
FT    DOMAIN     1492   1496       POLY-THR.
FT    DOMAIN     1559   1592       ASP/GLU-RICH (ACIDIC).
SQ    SEQUENCE   1596 AA;  167717 MW;  B944D6EF359D605 CRC64;
```

```
Query Match             4.7%;  Score 171.5;  DB 1;  Length 1596;
Best Local Similarity  18.1%;  Pred. No. 0.092;
Matches  161;  Conservative  93;  Mismatches  311;  Indels  325;  Gaps  35;

Qy     40  EQLSIPSQRFGDHGTMNSRSNNTSTLVHPGPSSOQPCGVERNLSVQHLDSSAANQATEKFV   99
                :  :||  || :     |||  | |||      ||||||:|||::
Db     166  QEETVLQKRFLE--SKNKRAAKKTDKKLPDPSQOHQQQQOHQQQQOHQQHQQAQTMLA  223

Qy    100  SQMSFMENVRSSAQHDQRKMVREED-----------FAVPVYINSRRSQSHGRTKSGIEKEKH  152
                 :    :| :: |||                   |   :|  |: |||   ||||| |
Db    224  GQL------QSSVHVQQKFLKRPAEDVDNGPDSFEPPHKLPNNNNSNSNNNNGNANANN  277

Qy    153  TPMVAPSSHHSIRFQBVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEE  212
                 :||           :|             :           |
Db    278  ---GGNGSNTGNNTNNNGNSTN--------------------NNGGSNNNGSENLTKPSVEIVQQ  319

Qy    213  IDLEKSA--SSHDRVND-------YNASLRQE--------------------SRNRLYRDGGKT  247
                 :|:  :    : |::|       |  :|                   ::|| ||| :
Db    320  LEFTTSPANSQPQISTNVTVKALTNTSVKSEPGVGGGGGGGGGNSGNNNNNGGGGGGG  379

Qy    248  RLKDTDNGAEBSHLATENHSQE-----------------------------GH  270
                    ||                                            ||
Db    380  NGNNNNNGGDHHQQQQHQHQQQQQQGGGLGGLGNNGRGGGPGGMATGPGGVAGGLGGM  439

Qy    271  GSPEDIDNDREYSKSRACASLQQINERASDDVSDDSMVDSISSIDVSPDDVVGILGQKRF  330
                ::: :     :|| ::|   :::||               :|||  :||  |::
Db    440  GMPPNMMSAQQKSALGNLANLVECKREPDHDFPD-----LGSLDKD----GGGGQ---  485
```

```
RN   [3]
RP   SEQUENCE OF 195-866 FROM N.A.
RC   TISSUE=Testis;
RA   Ottenwaelder B., Obermaier B., Mewes H.-W., Gassenhuber J.,
RA   Wiemann S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE OF 207-1157 FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=20039618; PubMed=10574461;
RA   Hirosawa M., Nagase T., Ishikawa K.-I., Kikuno R., Nomura N.,
RA   Ohara O.;
RT   "Characterization of cDNA clones selected by the GeneMark analysis
RT   from size-fractionated cDNA libraries from human brain.";
RL   DNA Res. 6:329-336(1999).
CC   -!- FUNCTION: May act to physically and functionally link
CC       transcription and pre-mRNA processing (By similarity).
CC   -!- SUBUNIT: Interacts with the repetitive C-terminal domain (CTD) of
CC       RNA polymerase II (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=O95104-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=O95104-2; Sequence=VSP_005879;
CC       Note=No experimental confirmation available;
CC   -!- SIMILARITY: Contains 1 RNA recognition motif (RRM) domain.
CC   -------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   -------------------------------------------------------------------------
DR   EMBL; AF023142; AAD09327.1; -.
DR   EMBL; AP001711; -; NOT_ANNOTATED_CDS.
DR   EMBL; AL117417; CAB59911.1; -.
DR   EMBL; AB032998; BAA86486.1; -.
DR   GK; O95104; -.
DR   InterPro; IPR000504; RNA_rec_mot.
DR   InterPro; IPR006569; RPR.
DR   Pfam; PF00076; rrm; 1.
DR   SMART; SM00582; RPR; 1.
DR   SMART; SM00360; RRM; 1.
DR   PROSITE; PS50102; RRM; 1.
DR   PROSITE; PS00030; RRM_RNP_1; FALSE_NEG.
KW   RNA-binding; Alternative splicing.
FT   NON_TER        1      1
FT   DOMAIN         1      8       POLY-ALA.
FT   DOMAIN       165    168       POLY-PRO.
FT   DOMAIN       314    321       POLY-ALA.
FT   DOMAIN       323    326       POLY-PRO.
FT   DOMAIN       518    592       RNA-BINDING (RRM).
FT   DOMAIN       726    733       POLY-PRO.
FT   DOMAIN       750    756       POLY-PRO.
FT   DOMAIN       967    976       POLY-GLN.
FT   VARSPLIC     776    798       STIAGINEDTTKDLSIGNPIPTV -> L (in isoform
FT                                 2).
FT                                 /FTId=VSP_005879.
FT   CONFLICT     270    274       MISSING (IN REF. 3).
FT   CONFLICT     293    296       T -> I (IN REF. 1).
FT   CONFLICT     296    296       T -> A (IN REF. 1).
FT   CONFLICT     309    309       T -> A (IN REF. 1).
FT   CONFLICT     312    312       V -> L (IN REF. 4).
FT   CONFLICT     562    562       H -> Y (IN REF. 1).
FT   CONFLICT     570    570       L -> V (IN REF. 1).
FT   CONFLICT     581    581       N -> T (IN REF. 2).
FT   CONFLICT     608    610       GVT -> CVI (IN REF. 1).
FT   CONFLICT     614    616       WDK -> CDN (IN REF. 1).
FT   CONFLICT     622    622       L -> M (IN REF. 1).
FT   CONFLICT     626    626       C -> Y (IN REF. 1).
```

```
FT   CONFLICT     818    818       AVPPAAPTNLPTPPVTQPV -> TCCTHSAHPSCNPACLP
FT                                 C (IN REF. 1).
FT   CONFLICT     836    837       SLLGTQGVAPGPVIGLQAPSTGLLGARPGL -> VSLLGKL
FT                                 FFFIISFPCPTPPHPTHPQVRFF (IN REF. 3).
FT   CONFLICT     854    854       A -> E (IN REF. 1).
FT   CONFLICT    1149   1149       S -> G (IN REF. 1).
SQ   SEQUENCE    1157 AA;  126846 MW;  176A77899BF01B CRC64;

Query Match               4.7%;  Score 171;  DB 1;  Length 1157;
Best Local Similarity    27.8%;  Pred. No. 0.067;
  Matches  71;  Conservative  19;  Mismatches  97;  Indels  68;  Gaps  11;

Qy    398  PLPHVVVKQRGDSEKTDQHKMESSA---------------ENVVGRLSNQG---  433
                            |                       ||  |
Db    794  PIPTVVSGARGNAESGDSVKMYGSAVPPAAPTNLPTPPVTQPVSLLGTQGVAPGPVIGLQ  853

Qy    434  --------HHQQSNYMPFANNPASPAPNGYCFPPQPP-PSGNHQQWLIPVMSPSEGLIYKP  486
Db    854  APSTGLLGARPGLIPL-QRPPGMPPPHLQRFPLMPPRPMPPIMMHRGPPPGPGGFAMPPP  912

Qy    487  HPGMAHTGHYGGYYGHYMP-TPMVMPQYHPGMGFPPPNGYFPPYGMMPTIMNPYCSSQQ  545
Db    913  H------------GMKGFFPPHGPFVRPGGMPGLGGPGPG-----------GP--EDR  947

Qy    546  QQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQQPTKSYP---RARKSRQGTG  602
Db    948  DGRQQPPQOQPQQPOPQAPQQPQQQQQ---QPPPSQQPPTQQQPQQFRNDNRQQFNSG  1004

Qy    603  SSPSGPQGISGSKSF  617
Db   1005  RD---QERFGRRSF  1015
```

```
RESULT 9
ID   YM11_YEAST     STANDARD;      PRT;    943 AA.
AC   P39523;
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 105.9 kDa protein in RPL15B-GCR3 intergenic region.
GN   YMR124W OR YM8564.06.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=97313268; PubMed=9169872;
RA   Bowman S., Churcher C.M., Badcock K., Brown D., Chillingworth T.,
RA   Connor R., Dedman K., Devlin K., Gentles S., Hamlin N., Hunt S.,
RA   Jagels K., Lye G., Moule S., Odell C., Pearson D., Rajandream M.A.,
RA   Rice P., Skelton J., Walsh S., Whitehead S., Barrell B.G.;
RT   "The nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   XIII.";
RL   Nature 387:90-93(1997).
RN   [2]
RP   SEQUENCE OF 770-943 FROM N.A.
RA   Pandit S., Sternglanz R.;
RL   Submitted (DEC-1992) to the EMBL/GenBank/DDBJ databases.
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; Z49273; CAA89273.1; -.
DR   EMBL; L07650; AAA35112.1; -.
DR   PIR; S54493; S54493.
```

```
DR   GermOnline; 142793; -.
DR   SGD; S0004731; YMR124W.
KW   Hypothetical protein.
FT   DOMAIN      59     66       POLY-GLN.
FT   DOMAIN      93    100       POLY-GLN.
FT   DOMAIN     164    169       POLY-GLY.
FT   DOMAIN     170    175       POLY-ASP.
FT   DOMAIN     371    381       POLY-GLN.
FT   DOMAIN     536    539       POLY-GLU.
FT   DOMAIN     715    718       POLY-SER.
SQ   SEQUENCE  943 AA;  105913 MW;  6631BEFBCD62E601 CRC64;

Query Match             4.6%;  Score 167;  DB 1;  Length 943;
Best Local Similarity  17.9%;  Pred. No. 0.086;
Matches 123;  Conservative  79;  Mismatches 191;  Indels 296;  Gaps  30;

Qy   134  SRRSQSHGRTK----------SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATC  183
                   ||||   |   ||              ::      :::       :|:|:
Db    10  TRNGQSSARVKLRNNLLNNDIGNIDIRDETP------------------ISRNGNDSNINIQPS  55

Qy   184  SKPEVRDQVKANARSGGFVISLDVSVTE-EIDLEKSASSHDRVNDYNASLRQESRNRL--  240
          : : ||  |          | ::| | |  || |:: || | |    |:       |:
Db    56  SVPQQQQQQQQYYRNGMNEAPIQAPLQQRQIPMQNVSQQQRQQQQVNFEYSNPHMNEIPL  115

Qy   241  ---YRDGGKTRLKDTDGAESHLATEN----------------HSQEGHGSPEIDNDR  280
                :      |: :    |||                     ::
Db   116  MQHNFTKPSLSNNRDNVNGKKASSFTQSSFSNFFKHKHQFGKSKKNTKGTGGGGDGDDDD  175

Qy   281  EYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGIIGQKRFWRARKAIANQ  340
             ||
Db   176  EV----------------------------------ILDDS-  182

Qy   341  QRVFAVQLFELHRILKVQKLIAASPDLLLDEISFIGKVSAKSY----------PVKKLLPSEF  393
              :|      ||||                   |: :::
Db   183  --------------ANSDLTFNDIQTPGHKGGDKYGVGGDSTPI--------  212

Qy   394  LVKPPLPHVVKQRGDSEKTDQHK----MESSAENVVGRLSNQGHHQQSNYMPFANNPPA  449
                  :|:|:            ::|
Db   213  ----IPTLIVTKDRGNMSNTEYRKYITNQRKTAMNAMAKQTKNG----------TLASLPPR  259

Qy   450  SPAPNGYCFPPQPPPSSGNHQQWLLPVM-SPSEGLIYKPHPGMAHTGHYGCYYGHYMPTPM  508
                   |||         ||:|:
Db   260  AMSLQSF-------PNGN--------PLMQAPT-------------------PHP-  287

Qy   509  VMPQYHPG---MGFPPPGNGYFFPYGMMPTTMNPYCSSQQQQQQQPNEQMNQFGHPGNLQ  565
                  : :                                       ||
Db   288  VSANYSRSNSLMSGPP----GQF------------RQPQQQQRMLPMNNYNNHPGQFQ  328

Qy   566  NT-------QQ--QQQR-----------SDNEPAP---------QQQQQPTKSYPRARK  595
                      ||                     ||||||
Db   329  NTPPVMPSGQQPPQQPRTLSLTNGPRYSPQNPRPFAGHQQISRQQQQQQQLHPMSEG  388

Qy   596  SRQGSTGSSPSGPQGISGSKSFRPFAAVDED-------------------SNIN  630
                    ::|                           |:
Db   389  YRITMSL-QSQNVPQG-----FNPWSPNDNDRKAVSMKQPISQSSISKNNSAYSIPNVQ  441

Qy   631  NAPEQTMTTTTTTTTRTTVVTQTTRDGG-------GVTRVIKV-VPH---  667
             :||                          |||:|:
Db   442  NNSLTTFSPSSPTDATAMPNSTKQGSSPLKKQVNIDQPIENKGKLNVLQLSTPQQNELKE  501

Qy   668  -NAKLAS-ENAARIFQSIQEERKRYDSSK  694
                |||:       |||:|| ::
Db   502  KERRKLAEMEKSLREREALVEEKEKERAEK  530
```

```
RESULT 10
EXPA_DROME
ID   EXPA_DROME   STANDARD;     PRT;  1429 AA.
AC   Q07436;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Expanded protein.
```

```
GN   EX.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Imaginal disks;
RX   MEDLINE=94094747; PubMed=8269855;
RA   Boedigheimer M., Laughon A.;
RT   "Expanded: a gene involved in the control of cell proliferation in
RT   imaginal discs.";
RL   Development 118:1291-1301(1993).
RN   [2]
RP   REVISIONS.
RA   Boedigheimer M.;
RL   Submitted (DEC-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Involved in the control of cell proliferation in
CC       imaginal disks. May bind to certain proteins of signal
CC       transduction pathways by interaction with their SH3 domains.
CC   -!- SUBCELLULAR LOCATION: Apical surface of disc cells.
CC   -!- MISCELLANEOUS: Mutations of expanded protein cause hyperplasmia of
CC       the imaginal disc resulting in wing overgrowth. This overgrowth is
CC       limited to specific regions along the 2 wing axes. Defects also in
CC       eyes, head, thorax and limbs where duplication and bulging often
CC       occur.
CC   -!- SIMILARITY: Contains 1 FERM domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L14768; AAB39774.1; -.
DR   PIR; T13720; T13720.
DR   FlyBase; FBgn0004583; ex.
DR   InterPro; IPR000299; Band_4.1.
DR   Pfam; PF00373; Band_41; 1.
DR   SMART; SM00295; B41; 1.
DR   PROSITE; PS00660; FERM_1; FALSE_NEG.
DR   PROSITE; PS00661; FERM_2; FALSE_NEG.
DR   PROSITE; PS50057; FERM_3; 1.
KW   Developmental protein; SH3-binding.
FT   DOMAIN      26     399       FERM.
FT   SITE      1008    1016       SH3-BINDING (POTENTIAL).
FT   SITE      1012    1020       SH3-BINDING (POTENTIAL).
FT   SITE      1149    1157       SH3-BINDING (POTENTIAL).
FT   DOMAIN     409    412       POLY-GLU.
FT   DOMAIN     782    788       POLY-PRO.
FT   DOMAIN     952    955       POLY-HIS.
FT   DOMAIN    1002   1005       POLY-PRO.
FT   DOMAIN    1011   1017       POLY-PRO.
FT   DOMAIN    1081   1084       POLY-PRO.
FT   DOMAIN    1149   1154       POLY-ALA.
FT   DOMAIN    1158   1168       POLY-ALA.
FT   DOMAIN    1170   1174       POLY-SER.
FT   DOMAIN    1199   1205       POLY-PRO.
FT   DOMAIN    1416   1424       POLY-GLN.
SQ   SEQUENCE  1429 AA;  153886 MW;  3CB08D2FC4862062 CRC64;

Query Match             4.5%;  Score 166.5;  DB 1;  Length 1429;
Best Local Similarity  19.2%;  Pred. No. 0.15;
Matches 167;  Conservative 100;  Mismatches 282;  Indels 323;  Gaps  39;

Qy     5  KDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSNNTST  64
             :||:|:|    |||:|:    |:     ||:|         ::|::
Db   408  REEEEAASQ--RLHASYACSRSLLPYKSKNE--QRISVISST----SSNTTSGIVSD  458

Qy    65  LVH--------------PGPSSQPCGVERNLSVQH-LDSSAANQATEKFVSQMSFMENVRSS  111
```

```
Db        |||  |||   |      |   ::  :| |   ||||
    459 RVHSEDELEIMINTPPAPLAAPS--TESLALAHLLDRPSVSRQTSS-VGQMS--------- 507

Qy  112 AQHDQRKMVREEDFAVPVYINSRRSQSHGR---TKSGIEK------EKHTPMVARPSS 160
          |:| :  :: :  ::  |   :   :|:|    ::|        :| ||||
Db  508 ------LKDLEEQLAALSVRPQDASSNGATIVTNSSVQRNSMGTTANDSSTATDSPSS 559

Qy  161 HHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVK-ANARSGGFVISLDVSVTEEIDL---- 215
          :|       |||| |     :::::::|
Db  560 QHN------TGSQ---CSSTCSTVVVTSPVNGAGASSSGAPIPVH-STSSSLELGFSH 607

Qy  216 --EKSASSHDRVNDY-NASLRQESRNRLYRDGGKTRLKDTDNGA---ESHLATENHSQEG 269
           ::: | :|:  :|: :   |      ::   ||
Db  608 TAQNSALSETSPDDFLSTSAREETES----VSGASGVYLAHGAPPTETSGVYTMHSSEL 663

Qy  270 HGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQXR 329
            :       :|     ||
Db  664 TGQSSEID-ESEKSSHYGMFQPQKLEE---THVSHSDSVDGKKKEDFRP---------- 708

Qy  330 FWRARKAIANQQRVFAVQLPELHRLHKLVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLL 389
          :|
Db  709 --RSDSNVSTGS------------------------SFRGDGSDPTDNKHSLL 735

Qy  390 PSEFLVKPPLPHVVVKQRG--DSEKTDQHKMESSAENVVCRLSNQGHHQQSNYMPFANNP 447
          :| ||       ||  :         : |      |       |||     ::
Db  736 SAEELTN-----LIIVGRGTYPSRKTVSSSLHSDCDYVTLPLGDQQEEE------VDQP 782

Qy  448 PASPAP------NGYCFPP------- 460
          |
Db  783 PAPPPPYSARHEKTGLCGPPIAKPIPKPIAVVAPKPDSPPCSPPVVPPAPIPAPPPAIRRR 842

Qy  461 QPPP----SGNHQQWLIPVMS-----PSEG------------LIIYKPHPG 489
          ||||
Db  843 DPPPYSISSKPRPTSLISVSSSAHPAPSAAGSMSSLKSEEVTARFITTRPQISILKAHTS 902

Qy  490 M-------------------------AHTGHYGG-----YYGHYMPT----- 506
          |
Db  903 LIPDGAKPSYAAPHHCSSVASSNGSVCSHQLSQQSLHNSNYAGGSQASLHHHVPSHHRH 962

Qy  507 ---------------------------PMVMPQYHPGMGFPPGPNGNGYFPPYGMMPTIM 537
                                      ::|: |   | ||      :|||
Db  963 SGSAAIGIVPYGLHKSTASLHHQQSCVLLPVIKPRQF--LAPPPPSLPRQPFPPPPN-- 1018

Qy  538 NPYCSCSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQOPTKSYPRARKSR 597
          :|:|  :     ::|      :  : :| : |||
Db  1019 HPHLASHIYEREMARKQLELY---------QQQLYSDVDVYLYPIQDPAVSQQEYLDAK 1068

Qy  598 QGS-----TGSSPSGPQGISGGSKSFRPFAAVDEDSNINNAPEQTMTTTTTRRTTVTQTT 652
          |||
Db  1069 QGSLLAAMAQAAPPPPH------HPYLAMQVSPAIYRSTPYLPLTLSTHSRYASTQNL 1120

Qy  653 RDGGGVTRVIKVVPHNAKLASENAARIFQSIQ 684
          :|
Db  1121 SD-----TYVQLPGPGYSPLYSPSMASLCSSYE 1148
```

RESULT 11

TF20_MOUSE

ID   TF20_MOUSE   STANDARD;   PRT; 1983 AA.
AC   Q9EPQ8; Q60792;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription factor 20 (Stromelysin 1 PDGF-responsive element-binding
DE   protein) (SPRE-binding protein) (Nuclear factor SPBP).
GN   TCF20 OR SPBP.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), CHARACTERIZATION, AND
RP   ALTERNATIVE SPLICING.

RC   STRAIN=BALB/c; TISSUE=Brain;
RX   MEDLINE=20568288; PubMed=10995766;
RA   Rekdal C., Sjoettem E., Johansen T.;
RT   "The nuclear factor SPBP contains different functional domains and
RT   stimulates the activity of various transcriptional activators.";
RL   J. Biol. Chem. 275:40288-40300(2000).
RN   [2]
RP   SEQUENCE OF 774-1965 FROM N.A. (ISOFORM 2).
RC   TISSUE=Fibroblast;
RX   MEDLINE=95280915; PubMed=7760812;
RA   Sanz L., Moscat J., Diaz-Meco M.T.;
RT   "Molecular characterization of a novel transcription factor that
RT   controls stromelysin expression.";
RL   Mol. Cell. Biol. 15:3164-3170(1995).
RN   [3]
RP   INTERACTION WITH JUN.
RX   MEDLINE=96279378; PubMed=8663478;
RA   Kirstein M., Sanz L., Moscat J., Diaz-Meco M.T., Saus J.;
RT   "Cross-talk between different enhancer elements during mitogenic
RT   induction of the human stromelysin-1 gene.";
RL   J. Biol. Chem. 271:18231-18236(1996).
RN   [4]
RP   INTERACTION WITH RNF4, TISSUE SPECIFICITY, AND MUTAGENESIS.
RX   MEDLINE=20408957; PubMed=10849925;
RA   Lyngsoe C., Bouteiller G., Damgaard C.K., Ryom D., Sanchez-Munoz S.,
RA   Noerby P.L., Bonven B.J., Joergensen P.;
RT   "Interaction between the transcription factor SPBP and the positive
RT   cofactor RNF4. An interplay between protein binding zinc fingers.";
RL   J. Biol. Chem. 275:26144-26149(2000).

CC   -!- FUNCTION: Transcriptional activator that binds to the regulatory
CC   region of MMP3 and thereby controls stromelysin expression. It
CC   stimulates the activity of various transcriptional activators such
CC   as JUN, SP1, PAX6 and ETS1, suggesting a function as a
CC   coactivator.
CC   -!- SUBUNIT: Homodimer (Probable). Interacts with RNF4 and JUN. Binds
CC   to the regulatory region of MMP3.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1;
CC   IsoId=Q9EPQ8-1; Sequence=Displayed;
CC   Name=2;
CC   IsoId=Q9EPQ8-2; Sequence=VSP_003986;
CC   -!- TISSUE SPECIFICITY: Expressed in brain, lung, liver, kidney and
CC   testes.
CC   -!- DEVELOPMENTAL STAGE: Isoform 2 is exclusively expressed at 7-11
CC   days of development. Isoform 1 is found only at low levels in 15-
CC   17 days embryos.
CC   -!- DOMAIN: The atypical PHD domain functions as a negative modulator
CC   of cofactor binding.
CC   -!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to
CC   frameshifts.
CC   --

DR   EMBL; AY007594; AAG28929.1; -.
DR   EMBL; U20282; AAA86495.1; ALT_SEQ.
DR   MGD; MGI:108399; Tcf20.
DR   GO; GO:0005634; C:nucleus; ISS.
DR   GO; GO:0003677; F:DNA binding; ISS.
DR   GO; GO:0003713; F:transcription co-activator activity; ISS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
DR   InterPro; IPR001965; Znf_PHD.
DR   SMART; SM00249; PHD; 1.
KW   Transcription regulation; Activator; DNA-binding; Zinc-finger;

```
Db   1421 PEGGTVATQEAEMEKRKCEVVSDLVSVTNQESNVEKPLPGPSEEWRGSGDDKVKTEAHVE 1480

Qy    417 ------------KMESSAENVVGRLSNQGHHQQS--NYMPFANNPPASPAPNGYCPPPQPPP 464
                          :                          :
Db   1481 TASTGKEPSGTMTSTASQKPG--GNQGRPDGSLGGAAPLIFPDSKNVAPVGILAPEANPK 1538

Qy    465 SGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTEMVMPQYHPGMGFPPPGN 524
                                                       |||
Db   1539 AEEKEN------------------------------DTVMISPKQE---SFPP-K 1559

Qy    525 GYFP----------PYGMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNE 577
               ||||                         ::|||||
Db   1560 GYFPSGKKKGRPIGSV--------NKQKKQQQQP----PPPQPQPMPEGSADGE 1602

Qy    578 PAPQQQQQPTK----SYPRARKSRQGSTGSSPSGPQ----------GISGSKSFRPF 620
               ::|:|:|          ||||:||:|::::        ::|:::||
Db   1603 PKPKKQRQRRERRKPGAQPRKKTKQAVPIVBPQEPEIKLIKYATQPLDKTDAKNKSFFPY 1662

Qy    621 AAVDEDSNIN----NAPEQTMT 638

Db   1663 IHVVNKCELGAVCTIINAEEEQT 1686
```

RESULT 12
```
ID   ADF1_CANAL      STANDARD;      PRT;    612 AA.
AC   P46589;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Adherence factor (Adhesion and aggregation mediating surface antigen).
GN   ADF1 OR AAF1.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE OF 35-612 FROM N.A.
RC   STRAIN=ATCC 36082;
RA   Edwards J.E.;
RL   Submitted (DEC-1994) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Jiang W., Finkler A., Koltin Y.;
RL   Submitted (JAN-1996) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: SURFACE ANTIGEN MEDIATING ADHESION AND AGGREGATION IN
CC       S.CEREVISIAE.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long  as its content is in no  way
CC   modified and this statement is not removed.  Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U18983; AAA62506.1; -.
DR   EMBL; U44747; AAA86758.1; -.
KW   Cell adhesion.
FT   DOMAIN        64     69       POLY-GLN.
FT   DOMAIN       100    103       POLY-SER.
FT   DOMAIN       117    134       POLY-GLN.
FT   DOMAIN       269    272       POLY-ASN.
FT   DOMAIN       517    520       POLY-PRO.
FT   DOMAIN       579    582       POLY-PRO.
FT   DOMAIN       596    605       POLY-GLN.
FT   CONFLICT     142    142       F -> N (IN REF. 1).
FT   CONFLICT     451    451       T -> I (IN REF. 1).
SQ   SEQUENCE    612 AA;   68794 MW;  77B60B45C35C1B23 CRC64;
```

Query Match   4.5%; Score 165; DB 1; Length 612;
Best Local Similarity 24.7%; Pred. No. 0.065;
Matches 76; Conservative 29; Mismatches 125; Indels 78; Gaps 12;

```
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN      1198   1219    LEUCINE-ZIPPER.
FT   DOMAIN      1282   1295    NUCLEAR LOCALIZATION SIGNAL.
FT   DNA_BIND    1565   1579    A.T HOOK.
FT   DOMAIN      1604   1628    NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      1812   1819    NUCLEAR LOCALIZATION SIGNAL.
FT   ZN_FING     1911   1962    PHD-TYPE (ATYPICAL).
FT   DOMAIN        62     69    POLY-ALA.
FT   DOMAIN       173    195    POLY-GLN.
FT   DOMAIN       249    277    SER-RICH.
FT   DOMAIN       322    350    POLY-GLN.
FT   DOMAIN      1585   1592    POLY-GLN.
FT   DOMAIN      1681   1684    POLY-PRO.
FT   DOMAIN      1793   1797    POLY-GLU.
FT   DOMAIN      1961   1983    POLY-GLU.
FT   VARSPLIC    1961   1983    CPLPPLQNKTAKGSLSTEQSERG  -> VRLWR (in
FT                              isoform 2)
FT                              /FTId=VSP_003986.
FT   MUTAGEN     1629   1629    A->T: LOSS OF INTERACTION WITH RNF4; WHEN
FT                              ASSOCIATED WITH S-1702, R-1736 AND V-
FT                              1737.
FT   MUTAGEN     1702   1702    P->S: LOSS OF INTERACTION WITH RNF4; WHEN
FT                              ASSOCIATED WITH T-1629, R-1736 AND V-
FT                              1737.
FT   MUTAGEN     1736   1737    CG-sRV: LOSS OF INTERACTION WITH RNF4;
FT                              WHEN ASSOCIATED WITH T-1629 AND S-1702.
FT   MUTAGEN     1926   1926    C->A: REDUCES THE INHIBITORY EFFECT OF
FT                              THE ATYPICAL PHD DOMAIN.
FT   MUTAGEN     1931   1931    C->A: REDUCES THE INHIBITORY EFFECT OF
FT                              THE ATYPICAL PHD DOMAIN.
FT   MUTAGEN     1936   1936    H->L: REDUCES THE INHIBITORY EFFECT OF
FT                              THE ATYPICAL PHD DOMAIN.
FT   MUTAGEN     1939   1939    C->A: REDUCES THE INHIBITORY EFFECT OF
FT                              THE ATYPICAL PHD DOMAIN.
SQ   SEQUENCE    1983 AA;  215284 MW;  E749A9355CB2971B CRC64;
```

Query Match   4.5%; Score 165.5; DB 1; Length 1983;
Best Local Similarity 19.4%; Pred. No. 0.25;
Matches 156; Conservative 96; Mismatches 265; Indels 287; Gaps 36;

```
Qy     19 HVNDADKGGPRAPPRNKMALY---EQLSIPSQ-----RFGDHGTMNSRSNN-------TSTL 65
             :| : : .:|||  : ::
Db    986 HETYRGNASPGAAAHDSISDYGPQDSRSTPMRRVPGRVGSRETMRGRSSSQYHDFAEKLK 1045

Qy     66 VHPGPSSQPCGVERNLSVQHLDSSAANQA----------TEKFVSQMSFMENVRSSAQ--- 113
             :|          ::|:
Db   1046 MSPGRSRGPGGDPHHMNPHMTFSERANRSSLHAFSPNSESLAS--AYHTNTRAHAYGDP 1103

Qy    114 -------HDQRKMVREEED-------------FAVPVY------INSRRSQSHGRTKSG 146
                                              :|
Db   1104 NTGLNSQLHYKRQMYQQQQEEYKDWASSSAQGVIAAAQHRQEGPRKSPRQQQFLDRVRSP 1163

Qy    147 IEKEKHTPMVAPS------------------------SHHSIRFQEVNQTGSKQN 177
             :: :|
Db   1164 LKNDKDGMMYGPPVGTYHDPSTQEAGRCLMSSDGLPAQSMELKHSSQKLQESRWDLSRQT 1223

Qy    178 VCLATCSKPEVRDQVKANARSG-----------GFVISLDVSVTEEIDL----EKSASSHD--- 223
                                                                |:
Db   1224 SPAKSSGPPGMSNQ------KRYGPPHEPDGHGLAESAQSKPSNVMLRLPGQEDHSSQNPL 1279

Qy    224 ---RVNDYNASL------RQESRNRLYRDGGKTRLKDTDNGAE----SHLATENHSQE 268
                                         :                   ||:
Db   1280 IMRRRVRSFISPIPSKRQSQDVKNSNADDKGRL-LHPSKEGADKAYNSYSHL---SHSQD 1335

Qy    269 GHGSPE-DIDNDREYSKSRAC--------ASLQQINEEASDDVSDDSMVDSISSIDVSPD 319
             |:|                          ||
Db   1336 IKSIPKRDSSKDLPNPDNRNCPAVTLTSPAKTLKLPPRKGRGLKLEAIVQKITSPNI--- 1392

Qy    320 DVVGILGCKRFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFL--- 375
                                                               ||
Db   1393 -------RRSASANSAE---------AGGDTVTLDDILSLKSGP 1420

Qy    376 ---GKVSAKSYPVKK------------LLPSEFIVKPPLPHVVVKQRG---DSEKTDQH-- 416
                :: :  :||              |||
```

```
Qy    375 LGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSA------ENVVGR 428
              ::|:  |  ||| :|:|                  |:|:|          : :|:
Db      1 MSSFNSNSNPANAQNLSSFQFQPPKEDVIF------------DTHQMSLSVPQSRYYASSMMQ 51

Qy    429 LSNQGH----HQQSNYMPFANN-PPA----SPAPNGYCFPP--QPPPSGNHQQWLIPV 475
              |||    ||     |  |  |||  |||
Db     52 QSQGGQPQSQTSQQQQQPFLMNIPPAFTQTQPQQMLYAMPPLQTQQPSSS--------- 102

Qy    476 MSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGRFPPPGNGYF----PPY 530
              |||    ||    |   :
Db    103 -SATTNNVVPPH----------HYNQQQSQQQQQQQYQQMQPQPNNMQFFDNTIPNY 150

Qy    531 GMMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSY 590
              |:  ||: :       |:|   :|          |
Db    151 LIMNQTISP---SQTQTTAQPNISYNY-----STQPQLQSAQPISHSQPQPQPQQATQ 200

Qy    591 PRARKSRQGSTGSGSSPSGPQGISG----------SKSFRPFAAVDEDSNINNAPEQT 636
              |||||||||||:| |||  | | ::  | :||:
Db    201 PRSNRSRSQTSFSKPRGSRQVSGSGSRSTGAKKQSAITSGSTGTGPARNADTGMTSVANST 260

Qy    637 MTTTTTTT 644
              |||| ||
Db    261 STTTMTTT 268
```

RESULT 13

```
PAB2_ARATH
ID    PAB2_ARATH      STANDARD;      PRT;   629 AA.
AC    P42731;
DT    01-NOV-1995 (Rel. 32, Created)
DT    01-NOV-1995 (Rel. 32, Last sequence update)
DT    10-OCT-2003 (Rel. 42, Last annotation update)
DE    Polyadenylate-binding protein 2 (Poly(A)-binding protein 2) (PABP 2).
GN    PAB2 OR AT4G34110 OR AT28A23.130.
OS    Arabidopsis thaliana (Mouse-ear cress).
OC    Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC    Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC    eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX    NCBI_TaxID=3702;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=cv. Columbia; TISSUE=Shoot;
RX    MEDLINE=94302150; PubMed=8029336;
RA    Hilson P., Carroll K.L., Masson P.H.;
RT    "Molecular characterization of PAB2, a member of the multigene family
RT    coding for poly(A)-binding proteins in Arabidopsis thaliana.";
RL    Plant Physiol. 103:525-533 (1993).
RN    [2]
RP    SEQUENCE FROM N.A.
RC    STRAIN=cv. Columbia;
RX    MEDLINE=20083488; PubMed=10617198;
RA    Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA    Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA    Harris B., Ansorge W., Brandt P., Grivell L.A., Rieger M.,
RA    Weichselgartner M., de Simone V., Obermaier B., Mache R., Mueller M.,
RA    Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidtheini T.,
RA    Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA    Vos P., Hoheisel J., Zimmermann W., Wedler H., Ridley P.,
RA    Langham S.-A., McCullagh B., Bilham L., Robben J., Vandenbussche F.,
RA    Braaken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA    Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA    Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA    Weitzenegger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA    Holzer E., Brandt A., Peters S., van Staveren M., Dirkse W.,
RA    Mooijman P., Klein Lankhorst R., Rose M., Hauf J., Koetter P.,
RA    Berneiser S., Hempel S., Feldpausch M., Lamberth S., Van den Daele H.,
RA    De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA    Van Montagu M., Rogers J., Cronin A., Quail M.A., Bray-Allen S.,
RA    Clark L., Doggett J., Hall S., Kay M., Lennard N., McLay K., Mayes R.,
RA    Pettett A., Rajandream M.A., Lyne M., Benes V., Rechmann S.,
RA    Borkova D., Bloecker H., Scharfe M., Grimm M., Loehnert T.-H.,
RA    Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA    Gabel C., Fuchs M., Fartmann B., Granderath K., Dauner D., Herzl A.,
```

```
RA    Neumann S., Argiriou A., Vitale D., Liguori R., Piravandi E.,
RA    Massenet O., Quigley F., Clabauld G., Muendlein A., Felber R.,
RA    Schnabl S., Hiller R., Schmidt W., Lecharny A., Aubourg S.,
RA    Chefdor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA    Gibbons T., Weber N., Vandenbol M., Bargues M., Terol J., Torres A.,
RA    Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA    Heinen L., Schwarz S., Scholler P., Heber S., Francs P., Bielke C.,
RA    Frishman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA    Zaccaria P., Bevan M., Wilson R.K., de la Bastide M., Habermann K.,
RA    Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,
RA    Sekhon M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
RA    Stoneking T., Kalicki J., Graves T., Harmon G., Edwards J.,
RA    Latreille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
RA    Minx P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
RA    Kramer J., Fulton L., Mardis E., Dante M., Pepin K., Hillier L.W.,
RA    Nelson J., Spieth J., Ryan E., Andrews S., Geisel C., Layman D.,
RA    Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshu C.,
RA    Antonoiu B., Zidanic M., Strong C., Sun H., Lamar B., Yordan C.,
RA    Ma P., Zhong J., Preston R., Vil D., Shekher M., Matero A., Shah R.,
RA    Swaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till J.,
RA    Granat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
RA    Chen E., Marra M.A., Martienssen R., McCombie W.R.;
RT    "Sequence and analysis of chromosome 4 of the plant Arabidopsis
RT    thaliana.";
RL    Nature 402:769-777(1999).
CC    -!- FUNCTION: Binds the poly(A) tail of mRNA.
CC    -!- TISSUE SPECIFICITY: Roots and shoots.
CC    -!- MISCELLANEOUS: A.thaliana contains at number of PABP genes which
CC        are expressed in an organ-specific manner.
CC    -!- SIMILARITY: Contains 4 RNA recognition motif (RRM) domains.
CC    --------------------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    --------------------------------------------------------------------------
DR    EMBL; L19418; AAA61780.1; -.
DR    EMBL; AL021961; CAA17561.1; -.
DR    EMBL; AL161584; CAB80128.1; -.
DR    PIR; T05425; T05425.
DR    HSSP; P11940; 1CVJ.
DR    InterPro; IPR002343; Hud_Sxl_RNA.
DR    InterPro; IPR002004; PABP/HECT.
DR    InterPro; IPR006515; PARP_1234.
DR    InterPro; IPR000504; RNA_rec_mot.
DR    Pfam; PF00658; PABP; 1.
DR    Pfam; PF00076; rrm; 4.
DR    PRINTS; PR00961; HUDSXLRNA.
DR    SMART; SM00517; PolyA; 1.
DR    SMART; SM00360; RRM; 4.
DR    TIGRFAMs; TIGR01628; PABP-1234; 1.
DR    PROSITE; PS00030; RRM_RNP_1; 3.
DR    PROSITE; PS50102; RRM; 4.
KW    RNA-binding; Nuclear protein; Repeat; Multigene family.
FT    DOMAIN      36    114   RNA-BINDING (RRM) 1.
FT    DOMAIN     124    196   RNA-BINDING (RRM) 2.
FT    DOMAIN     215    292   RNA-BINDING (RRM) 3.
FT    DOMAIN     318    395   RNA-BINDING (RRM) 4.
SQ    SEQUENCE   629 AA;  68672 MW;  0F809818D08BDC7E CRC64;
```

```
Query Match            4.5%;  Score 164.5;  DB 1;  Length 629;
Best Local Similarity 20.7%;  Pred. No. 0.072;
Matches 126;  Conservative 71;  Mismatches 210;  Indels 203;  Gaps  30;

Qy    67 HPGPSSQPCGVERNLSVQHLDSS------AANQATEKFVSQMSFMENVRSSAQHDQRKMVRE 122
             | |||:   ||  ||::  |||::      |:
Db    113 HRDPSVRRSGA-GNIFIKNLDESIDHKALHDTFSSFGNIVSCKVAVDSSGQ-------- 162

Qy   123 EEDFAVPYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLAT 182
             |:   |:    |:       ::|| |:|:  |:  :|  |:   |:
```

```
Db   163  SKGYGFVQYANEESAQKAIEKLNGMLLNDKQVYVGP----FLRRQERDSTANKTKF----  214
              :|| ::          ::  :|:|  ::||
Qy   183  CSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYR  242
          :     |: :           :||:
Db   215  -----------TNVYVKNLAESTTDD-DLKNAFGEYGKITSAVVMKDGEGKSKGF-     257
                                            ||:|:    |:
Qy   243  DGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINE-----  296
           :   |:     ||            |  :||:|
Db   258  --GFVNFENADDAAR---AVE--SLNGHKF----DDKEWYVGRAQKKSERETELRVRYE  305
           :   :|             ||
Qy   297  ----EASDDVSDDSMVDSISSIDVSPDD-------VVGILGQKRFWRARKAIANQQRVF  344
              :||   | ::    :|:|
Db   306  QNLKEAADKFQSSNLY--VKNLDPSISDEKLKEIRSPFGTVTSSKVMRDPNGTSKGSGFV  363
               |:|:|  |:       |:          :          |:
Qy   345  AVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVV  404
              |:                              :|:
Db   364  AF-------------ATPEEATEAMSQLSGKMIESKPL------------YVAI       392
            :
Qy   405  KQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPP  464
              :     |::                    |:        ::
Db   393  AQRKEDRRV----------RL----QAQFSQVRPVAMQPSVGPR----MPVYPP       428
              ||:                |:            |:||
Qy   465  SGNHQQWLIPVMSPSEGLIYKPHPGMAHTCHYGGVYGHYMPTPMVMPQYHPGMGFPP---  521
           ::             |:||          |||
Db   429  GG-----------------------PGIGQMFYG-----QAPPAMIPPQ--PGYGYQQQLV  460
                                   |:|||
Qy   522  ---PGNGYFPPYGMGMPTIMNPYCSSQQQQQQQPNEQMNQFGHPDGNLQNTQQQQQRSDNE  577
              ||   |   |                               :|:|
Db   461  PGMRPGGGPVPSF-FMPMV-------QPQQQRPGGGR---RPCGIQHSQQQN-------    501
             |:                              |||:
Qy   578  PAPQQQQQPTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDE--------DSNI  629
              ||                           |:|||
Db   502  --PMMQQQ---MHPRGRMFRY-----------------PQGRGGSGDVPPYDMGNMMPLTIGALASNL  547
              ||||          |:|||

Qy   630  NNA-PEQTMT  638
          :|| |||
Db   548  SNATPEQQRT  557
```

RESULT 14
OSA_DROME
ID   OSA_DROME   STANDARD;   PRT;   2716 AA.
AC   Q8IN94; O61603; Q9VEG7;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Trithorax group protein OSA (Eyelid protein).
GN   OSA OR ELD OR CG7467.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A., FUNCTION, SUBCELLULAR LOCATION, TISSUE
RP   SPECIFICITY, AND DEVELOPMENTAL STAGE.
RX   MEDLINE=97415319; PubMed=9271118;
RA   Treisman J.E., Luk A., Rubin G.M., Heberlein U.;
RT   "eyelid antagonizes wingless signaling during Drosophila development
RT   and has homology to the Bright family of DNA-binding proteins.";
RL   Genes Dev. 11:1949-1962 (1997).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195 (2000).
RN   [3]
RP   FUNCTION, AND DEVELOPMENTAL STAGE.
RX   MEDLINE=99112962; PubMed=9895321;
RA   Vazquez M., Moore L., Kennison J.A.;
RT   "The trithorax group gene osa encodes an ARID-domain protein that
RT   genetically interacts with the brahma chromatin-remodeling factor to
RT   regulate transcription.";
RL   Development 126:733-742(1999).
RN   [4]
RP   DNA-BINDING, AND IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM AND SNR1.
RX   MEDLINE=20069333; PubMed=10601025;
RA   Collins R.T., Furukawa T., Tanese N., Treisman J.E.;
RT   "Osa associates with the Brahma chromatin remodeling complex and
RT   promotes the activation of some target genes.";
RL   EMBO J. 18:7029-7040(1999).
RN   [5]
RP   FUNCTION.
RX   MEDLINE=99403006; PubMed=10471712;
RA   Staehling-Hampton K., Ciampa P.J., Brook A., Dyson N.;
RT   "A genetic screen for modifiers of E2F in Drosophila melanogaster.";
RL   Genetics 153:275-287(1999).
RN   [6]
RP   IDENTIFICATION IN A BRAHMA COMPLEX WITH BRM; OSA; MOR; SNR1; DALAO;
RP   BAP55; BAP60 AND BAP47, AND FUNCTION AS COACTIVATOR.
RX   MEDLINE=20270023; PubMed=10809665;
RA   Kal A.J., Mahmoudi T., Zak N.B., Verrijzer C.P.;
RT   "The Drosophila brahma complex is an essential coactivator for the
RT   trithorax group protein zeste.";
RL   Genes Dev. 14:1058-1071(2000).
RN   [7]
RP   FUNCTION AS A COREPRESSOR.
RX   MEDLINE=20573925; PubMed=11124806;
RA   Collins R.T., Treisman J.E.;
RT   "Osa-containing Brahma chromatin remodeling complexes are required
RT   for the repression of wingless target genes.";
RL   Genes Dev. 14:3140-3152(2000).
RN   [8]
RP   FUNCTION AS A COREPRESSOR, AND INTERACTION WITH PNR AND CHI.
RX   MEDLINE=22515897; PubMed=12629041;
RA   Heitzler P., Vanolst L., Biryukova I., Ramain P.;
RT   "Enhancer-promoter communication mediated by Chip during
RT   Pannier-driven proneural patterning is regulated by Osa.";
RL   Genes Dev. 17:591-596(2003).

```
FT   CONFLICT     2637   2637       G -> E (IN REF. 1).
SQ   SEQUENCE     2716 AA; 284063 MW; EFAE76CB51C7C675 CRC64;

     Query Match             4.5%;  Score 164.5;  DB 1;  Length 2716;
     Best Local Similarity  24.7%;  Pred. No. 0.41;
     Matches    72; Conservative  13; Mismatches 118; Indels 89; Gaps 12;

Qy        397 PPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQ-----------QSN 439
              |||     ||||| |
Db        142 PPHPHPAYGRYHADPNMDPYRYGGPLPGGKPPQQQOPHPQQQPGPGGSPNRPPQQR 201

Qy        440 YMPFANNPPASPAP--NGYCFPPQPPPSGNH-----------------QQWLIPV 475
              ||:|                             |:|
Db        202 YIP--GQPPQGPTPTLNSLLQSSNPPPPQQHRYANTYDPQQAAAAAAAQQQQAGGP 259

Qy        476 MSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPM-VMPQYHPGMGF--PPPGN----GY 526
              ||:|                             |:|
Db        260 PPPGHG---PPPPQHQPSPYGGQQGWAPPPRPYSPQLGPSQQYRTPPTNTSRGQSPY 315

Qy        527 FPPYGVMGMPTIMNPYCSSQQQQQQQQQPNEQMNQFGH-------------PGNLQNTQQQQQRSD 575
              ||:|
Db        316 PPAHGQN--SGSYPSSPQQQQQQQQQQQQAGQPGGPVPGGPPPGTGQPPQQNTPPT 372

Qy        576 NEPAPQQQQQP-----------------TKSYPRARKSRQGSTGSSPSGPQG 610
              ::|:|
Db        373 SQYSPYPQRYPTPPGLPAGGSNHRTAYSTHQYPEPNRPWPG--GSSPSPSGG 422
```

```
RESULT 15
NCO2_XENLA
ID   NCO2_XENLA     STANDARD;     PRT;  1516 AA.
AC   Q9W705;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Nuclear receptor coactivator 2 (NCoA-2) (Transcriptional intermediary
DE   factor 2) (XTIF2).
GN   NCOA2 OR TIF2.
OS   Xenopus laevis (African clawed frog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Amphibia; Batrachia; Anura; Mesobatrachia; Pipoidea; Pipidae;
OC   Xenopodinae; Xenopus.
OX   NCBI_TaxID=8355;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20171035; PubMed=10704837;
RA   de la Calle-Mustienes E., Gomez-Skarmeta J.L.;
RT   "XTIF2, a Xenopus homologue of the human transcription intermediary
RT   factor, is required for a nuclear receptor pathway that also
RT   interacts with CBP to suppress Brachyury and XMyoD.";
RL   Mech. Dev. 91:119-129(2000).
CC   -!- FUNCTION: Transcriptional coactivator for steroid receptors and
CC       nuclear receptors. Coactivator of the steroid binding domain
CC       (AF-2) but not of the modulating N-terminal domain (AF-1).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DEVELOPMENTAL STAGE: Expressed homogeneously during late blastula-
CC       early gastrula stage and later becomes highly expressed in the
CC       notochord.
CC   -!- SIMILARITY: Contains 1 basic helix-loop-helix (bHLH) domain.
CC   -!- SIMILARITY: Contains 1 PAS (PER-ARNT-SIM) dimerization domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AJ243119; CAB45389.1; -.
DR   InterPro; IPR001092; HLH_basic.
DR   InterPro; IPR000014; PAS_domain.
DR   InterPro; IPR008955; Src-1.
```

```
CC   -!- FUNCTION: Trithorax group (trxG) protein required for embryonic
CC       segmentation, development of the notum and wing margin, and
CC       photoreceptor differentiation. Required for the activation of
CC       genes such as Antp, Ubx and Eve. Binds to DNA without specific
CC       affinity, suggesting that it is recruited to promoters by
CC       promoter-specific proteins. Essential component of the Brahma
CC       complex, a multiprotein complex which is the equivalent of the
CC       yeast SWI/SNF complex and acts by remodelling the chromatin by
CC       catalyzing an ATP-dependent alteration in the structure of
CC       nucleosomal DNA. This complex can both serve as a transcriptional
CC       coactivator or corepressor, depending on the context. Acts as an
CC       essential coactivator for Zeste, which recruits the whole complex
CC       to specific genes. In contrast, it acts as a corepressor for Wg
CC       target genes, possibly via an interaction with Pan and Gro. It
CC       also acts as a negative regulator for proneural achaete-scute,
CC       when it is directly recruited by Pan and Chi. Also represses E2f
CC       activation.
CC   -!- SUBUNIT: Component of the Brahma complex, which is composed of
CC       Brm, Osa, Mor, Snr1/Bap45, Bap111/Dalao, Bap55, Bap60 and Bap47.
CC       Interacts with Pnr and Chi via its EHD domain.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- TISSUE SPECIFICITY: Ubiquitously expressed in early embryo. In
CC       third instar larvae, it is ubiquitously expressed in wing and eye-
CC       antenna imaginal disks, with a stronger expression in a band just
CC       anterior to the morphogenetic furrow.
CC   -!- DEVELOPMENTAL STAGE: Expressed both maternally and zygotically.
CC   -!- DOMAIN: The ARID domains mediates the binding to DNA.
CC   -!- SIMILARITY: Contains 1 ARID domain.
CC   -!- SIMILARITY: Contains 1 EHD (Eyelid homology) domain.
CC   -!- CAUTION: Ref.2 (AAF55457) sequence differs from that shown due to
CC       erroneous gene model prediction.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF053091; AAC06254.1; -.
DR   EMBL; AE003718; AAF55457.1; ALT_SEQ.
DR   EMBL; AE003718; AAN13750.1; -.
DR   PIR; T13049; T13049.
DR   FlyBase; FBgn0003013; osa.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0003677; F:DNA binding; IDA.
DR   GO; GO:0046530; P:photoreceptor cell differentiation; IMP.
DR   GO; GO:0045449; P:regulation of transcription; IDA.
DR   GO; GO:0007379; P:segment specification; IMP.
DR   GO; GO:0008587; P:wing margin morphogenesis; IMP.
DR   GO; GO:0016055; P:Wnt receptor signaling pathway; IMP.
DR   InterPro; IPR001606; ARID.
DR   InterPro; IPR008938; ARM.
DR   InterPro; IPR006031; XYPPX.
DR   Pfam; PF01388; ARID; 1.
DR   Pfam; PF02162; XYPPX; 8.
DR   SMART; SM00501; BRIGHT; 1.
KW   Chromatin regulator; Nuclear protein; Developmental protein.
KW   Transcription regulation; DNA-binding; Activator; Repressor;
FT   DOMAIN        997   1111       ARID.
FT   DOMAIN       1769   2517       EHD.
FT   DOMAIN         19    380       PRO-RICH.
FT   DOMAIN        174    873       GLN-RICH.
FT   DOMAIN        619   1222       GLY-RICH.
FT   DOMAIN       1271   1751       GLN-RICH.
FT   DOMAIN       1730   1745       HIS-RICH.
FT   DOMAIN       2589   2624       SER-RICH.
FT   DOMAIN       2625   2716       ALA-RICH.
FT   CONFLICT       61     61       MISSING (IN REF. 1).
FT   CONFLICT     1169   1169       V -> G (IN REF. 1).
FT   CONFLICT     1795   1795       M -> T (IN REF. 1).
```

```
DR   Pfam; PF00989; PAS; 1.
DR   SMART; SM00353; HLH; 1.
DR   SMART; SM00091; PAS; 1.
DR   PROSITE; PS50888; HLH; 1.
DR   PROSITE; PS50112; PAS; 1.
KW   Transcription regulation; Activator; Nuclear protein.
FT   DOMAIN      116    180       PAS.
FT   DOMAIN     1237   1273       POLY-GLN.
SQ   SEQUENCE   1516 AA; 166156 MW; 0985lC00AB439A4A CRC64;

Query Match             4.5%;  Score 163.5;  DB 1;  Length 1516;
Best Local Similarity  20.5%;  Pred. No. 0.23;
Matches 162;  Conservative  79;  Mismatches 283;  Indels 265;  Gaps 37;

Qy      4 GKDEEKILEPMFPRLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGD--HGTMNSRSN 60
            :| :: || |    |                      ::|  |    | |
Db    621 GKGQQKLL-----KLLTTKSDQ----------------MEPSTLPSNTLGDMNKDSLSNFASN 662

Qy     61 NTSTLVHPGPSSQPCGVERNLSVQHL-DSSA---------ANQATEKFVSQMSFMENVRSS 111
            :   |  |  |              ||                |  |: :|||  |::
Db    663 SMSASAH-GTSLK---EKHKILHRLLQDSSSPVDLAKLTAEATGKELSQES--------NS 711

Qy    112 AQHDQRKMVREEEDFAVPVVINSRRSQSHGRTKSG1EKEKHTPMVAPSSHHSIRFQEVNQ 171
            :|:    ||:  :     |                |: ||   :|  ||
Db    712 TGPGSEVT1KQE------PV---SPKKKEHALLRYLLDKDDTTDNVADITPKLER------ 757

Qy    172 TGSKQNVCLATCSKPEVRDQVKANARSGGF----VISLDVSVTEEI--DLEKSASSH--- 222
             ::: | |        |                          |      | :
Db    758 --ADNKVDFSSCPK--LSAVKAEKEEPNFGHTDQPGSDFDNLDEILDDLQNSQLSQLFS 812

Qy    223 DRVNDYNASLRQESRNRLYRDGG----------KTRL--KDTDNGAESHLATENHSQEG 269
             | : | | |||:|| |  |             ::|    |  : :|  ||   |:|
Db    813 DTRHDGNSADKQA1MNDLMQLAGENSTGLPAWAQKQRMLRMQQNNGFNSQLA----AQLG 868

Qy    270 HGSPEDIDNDREYSKSRACASLQQINEEA------------SDDVSDDSMVDSISSIDVSPDDVV 322
            |:    :  |:: |: :                 :: | :  | :||   | :|
Db    869 RLPNQNLPLDIHFQSQASAGSFAQMRSSGPYTTVPQSGVINNQAMMGSQGNVRNSSPGIV 928

Qy    323 GILGQK-----RFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGK 377
            |:||       :: |  | :  | :    | :  |: |          |  ||:
Db    929 GVNGPRPPLKPGDWGSQASAVRPACPTTSTAMNRHDM------SRSP------------T 970

Qy    378 VSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSE----------KTDQHKMESSAERNVVGRLSN- 431
            ||   |       |  :| | :| |  |: :              ::
Db    971 ASIPMRPGSGQVCPRQVL-----QSAVMNMGGSBELDMN1SGPQYTQQQAPPNQTAPWPNR 1024

Qy    432 -----QGHHQQSNVMPFANNPPASPAPNGYCFP--PQPPPSGNHQQWL- 472
                  |:|: ||     |  ||: :|             | ||::|
Db   1025 ILLTEQPSFNNQNRQPF----GSPADDLICQPIVSESPADDGNLLDQLYMALRNFDGLE 1079

Qy    473 -------IPVM-----------SPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQ 512
                    ::             :: |  |       | |   :|
Db   1080 EIDRALGIPEMVSQGQAVEQESFGSPESNLMMEQKPPV-------YNHAYANQGQMAQ 1130

Qy    513 --YHP--GMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQPNEQMNQFGH------ 560
             | :      |          |:  |: ||| ||||||||  |: :||
Db   1131 NSYQPMQDPGFNFMGQR--PSYGILRMQNRPGLRPTGMVQNQPNQLRLQLQHRLQAQNRQ 1188

Qy    561 ------------PGNL-------------------------------------------Q 565
                          |:|
Db   1189 QLMNPINNVSNMNLAMRPGVPGQLREQGPINAQMLAQRQREILSQHLRQKQLQQQQQQQ 1248

Qy    566 NTQQQQQRSDNEPAPQQQQQQQPTKSYPRARKSRQ-----GSTG----SSPSGPQG 610
            | |||||   | |        || | | || |       |:|      ||  |||
Db   1249 QQQQQQQQQQQQQQQQQQHRAMMMRSQGLAMPPNMVGSSGIPASINSPRIPQG 1308

Qy    611 ISGSKSFRP 619
            ::| :: |
Db   1309 STQQFPFPP 1317
```

RESULT 16

CPH1_CANAL

```
ID   CPH1_CANAL     STANDARD;      PRT;    656 AA.
AC   P43079;
DT   01-NOV-1995 (Rel. 32, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Transcription factor CPH1.
GN   CPH1 OR ACPR.
OS   Candida albicans (Yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX   NCBI_TaxID=5476;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=SC5314;
RX   MEDLINE=94364983; PubMed=8083193;
RA   Malathi K., Ganesan K., Datta A.;
RT   "Identification of a putative transcription factor in Candida
RT   albicans that can complement the mating defect of Saccharomyces
RT   cerevisiae stel2 mutants.";
RL   J. Biol. Chem. 269:22945-22951(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95084165; PubMed=7992058;
RA   Liu H., Koehler J., Fink G.R.;
RT   "Suppression of hyphal formation in Candida albicans by mutation of a
RT   STE12 homolog.";
RL   Science 266:1723-1726(1994).
RN   [3]
RP   ERRATUM.
RA   Liu H., Koehler J., Fink G.R.;
RL   Science 267:17-17(1994).
RN   [4]
RP   CHARACTERIZATION.
RX   MEDLINE=95100929; PubMed=7802635;
RA   Singh P., Ganesan K., Malathi K., Ghosh D., Datta A.;
RT   "ACPR, a STE12 homologue from Candida albicans, is a strong inducer
RT   of pseudohyphae in Saccharomyces cerevisiae haploids and diploids.";
RL   Biochem. Biophys. Res. Commun. 205:1079-1085(1994).
CC   -!- FUNCTION: Transcription factor involved in the formation of
CC       pseudohyphae and hyphae. It is likely to play a role in the
CC       developmental switch between yeast and mycelial forms. May be
CC       involved in a signal transduction system, strengthening the
CC       possibility of a sexual phase up to now undetected, and similar to
CC       that of the yeast mating pathway.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: BELONGS TO THE STE12 TRANSCRIPTION FACTOR FAMILY.
CC
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; L16451; AAA59355.1; ALT_TERM.
DR   EMBL; U15152; AAA64692.1; -.
DR   PIR; A54767; A54767.
DR   InterPro; IPR003120; TF_STE.
DR   Pfam; PF02200; STE; 1.
DR   SMART; SM00424; STE; 1.
KW   Transcription regulation; DNA-binding; Activator; Nuclear protein.
FT   DNA_BIND     44    154       BY SIMILARITY.
FT   DOMAIN      455    458       POLY-ALA.
FT   DOMAIN      555    563       POLY-GLN.
FT   DOMAIN      574    586       POLY-GLN.
FT   DOMAIN      593    596       POLY-LYS.
FT   CONFLICT    482    482       I -> M (IN REF. 2).
FT   CONFLICT    510    510       A -> T (IN REF. 2).
FT   CONFLICT    610    610       K -> N (IN REF. 2).
SQ   SEQUENCE   656 AA;  72978 MW;  B55594BF019A9B09 CRC64;

Query Match             4.4%;  Score 162.5;  DB 1;  Length 656;
```

Best Local Similarity 20.5%; Pred. No. 0.096; Gaps 25;
Matches 118; Conservative 68; Mismatches 204; Indels 185;

```
Qy   176  QNVCLATCSKPEV-------RDQVKANA-----------------RSGGFVISLDVSV 209
               :||:|:|:|    :::|:|:|                        :|:::|:|
Db   128  KNSCLRTQKKQKVFFWFNVPHDKLMADALERDLKKEKMGQRPTTMAHREPALSPHYDESS 187

Qy   210  TEEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAEBSHLATENH--- 265
               :|:||                   :|:|:|    :|:|:|:|
Db   188  SLYTQLGKHMETQKRINDAATSSTSNTATTLTDTGVSSGLNNTTSGGGSDSATSTHNNNE 247

Qy   266  --SQEGHGSPEDIDNDREY-SKSRACASLQQINE---------EASDDVSDDSMVDSISSIDV 316
            :: .:||: .:||:|                    :|                ||:
Db   248  ASTKPSNGSEK---SSPEYTTARGRDEFGFLNEATPSQYKANSDYEDDFPLDYINQTTQ 304

Qy   317  SPDDVVGIIGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFL- 375
               :|:|:||                        ||:||                 :|:
Db   305  NSEDYITL----------------DANYQAG-----------SYANMIEDNYDSFLDATLFIP 340

Qy   376  ------GKVSAKSYPVKKLLPSEFL-------VKPPLPHVVVKQRGDSEKTDQHKMESSA 422
               :::||:|                       :|:|||
Db   341  PSLGVPTGTAATATTSNQVAFNDEYLIEQAQPIRTPLPPI-----------------SS 382

Qy   423  ENVVGRL------------SNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNH-QQWL 472
                |||                      ||:    :|:|
Db   383  STISGLLQPKSAAKFSLQSANGGBEFFPAVQNDP-STANAGFV---PPISAKYATQFA 437

Qy   473  IPVMSPSEGLIYKPHPGMAHTGHYGG------YYGH----YMPT-PMVMPQYHP------ 515
               :|:    :|:                  ||:|    |:|    |::
Db   438  TRQVATPTYIKAIPQTGAAAATGNGGQPQQYYDQATGNAFYPAEIPVLYNVVHPESEYWT 497

Qy   516  -----------------GMGPPPGNGYF-------PPY-----GMMPTIMNPYCSS 543
                                |||:||:|        |:|        |:|:|
Db   498  NNSGAVATTAAAAAPMYDASGFPIPINQSYMVMNEHEMVPYQYMNSNGAMIGMIPPH--- 554

Qy   544  QQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGS 603
               ||||||
Db   555  QQQQQQQQ---QQIAMGYQSMLRQQQQQQQ---------QQQQQPSSTMTKKKKQIHSFNNN 604

Qy   604  SPSGPQGISGGSKSFRPFAAVDEDSNINNAPEQTMT 638
               ||||||          ::|:|:|
Db   605  KSLSSKG-----------GGITKKSHDNNNHSKVKT 629
```

RESULT 17
ATX1_HUMAN
ID   ATX1_HUMAN   STANDARD;   PRT;   816 AA.
AC   P54253; Q9UJG2; Q9Y4J1;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Ataxin-1 (Spinocerebellar ataxia type 1 protein).
GN   SCA1 OR ATX1.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Cerebellum;
RX   MEDLINE=95038838; PubMed=7951322;
RA   Banfi S., Servadio A., Chung M.-Y., Kwiatkowski T.J. Jr., McCall A.E.,
RA   Duvick L.A., Shen Y., Roth E.J., Orr H.T., Zoghbi H.Y.;
RT   "Identification and characterization of the gene causing type 1
RT   spinocerebellar ataxia.";
RL   Nat. Genet. 7:513-519(1994).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Tubby B.;
RL   Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE OF 189-231 FROM N.A.
RX   MEDLINE=96177682; PubMed=8634720;

RA   Quan F., Janas J., Popovich B.W.;
RT   "A novel CAG repeat configuration in the SCA1 gene: implications for
RT   the molecular diagnostics of spinocerebellar ataxia type 1.";
RL   Hum. Mol. Genet. 4:2411-2413(1995).
RN   [4]
RP   SELF-ASSOCIATION SITE.
RX   MEDLINE=97252384; PubMed=9097953;
RA   Burright E.N., Davidson J.D., Duvick L.A., Koshy B., Zoghbi H.Y.,
RA   Orr H.T.;
RT   "Identification of a self-association region within the SCA1 gene
RT   product, ataxin-1.";
RL   Hum. Mol. Genet. 6:513-518(1997).
RN   [5]
RP   RNA-BINDING DOMAIN.
RX   MEDLINE=21065637; PubMed=11136710;
RA   Yue S., Serra H.G., Zoghbi H.Y., Orr H.T.;
RT   "The spinocerebellar ataxia type 1 protein, ataxin-1, has RNA-binding
RT   activity that is inversely affected by the length of its polyglutamine
RT   tract.";
RL   Hum. Mol. Genet. 10:25-30(2001).
CC   -!- FUNCTION: Binds RNA in vitro. May be involved in RNA metabolism.
CC   The expansion of the polyglutamine tract may alter this function.
CC   -!- SUBUNIT: Interacts with LANP and A1U.
CC   -!- SUBCELLULAR LOCATION: Nuclear and cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=1;
CC   Comment=At least 2 isoforms are produced;
CC   Name=1;
CC   IsoId=P54253-1; Sequence=Displayed;
CC   -!- TISSUE SPECIFICITY: Widely expressed throughout the body.
CC   -!- POLYMORPHISM: The poly-Gln region of SCA1 is highly polymorphic (4
CC   to 39 repeats) in the normal population and is expanded to about
CC   40-83 repeats in sca1 patients. Longer expansions result in
CC   earlier onset and more severe clinical manifestations of the
CC   disease.
CC   -!- DISEASE: Defects in SCA1 are the cause of spinocerebellar ataxia
CC   type 1 (SCA1) [MIM:164400]; also known as olivopontocerebellar
CC   atrophy I (OPCA I). SCA1 is an autosomal dominant
CC   neurodegenerative disorder characterized by progressive neuronal
CC   loss in the cerebellum, brain stem and spinocerebellar tracts.
CC   Clinical features are cerebellar ataxia, dysarthria,
CC   ophtalmoparesis, muscle wasting and neuropathy. Onset of the
CC   disease usually occurs in the third or fourth decade of life and
CC   death occurs ten to twenty years later.
CC   -!- MISCELLANEOUS: The self-association seems to be necessary to form
CC   nuclear aggregates.
CC   --
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   used for commercial purposes and this statement is not removed. Usage by and for commercial
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --
DR   EMBL; X79204; CAA55793.1; -.
DR   EMBL; AL009031; CAA15622.1; -.
DR   EMBL; S82497; AAD14401.1; -.
DR   PIR; S46268; S46268.
DR   Genew; HGNC:10548; SCA1.
DR   MIM; 601556; -.
DR   MIM; 164400; -.
DR   GO; GO:0005737; C:cytoplasm; TAS.
DR   GO; GO:0005634; C:nucleus; TAS.
DR   InterPro; IPR003652; Ataxin.
DR   SMART; SM00536; AXH; 1.
KW   RNA-binding; Nuclear protein; Polymorphism; Triplet repeat expansion;
KW   Alternative splicing.
FT   DOMAIN   197   226   POLY-GLN.
FT   DOMAIN   495   605   SELF-ASSOCIATION SITE.
FT   DOMAIN   541   767   RNA BINDING.
FT   DOMAIN   795   798   NUCLEAR LOCALIZATION SIGNAL
FT   (BY SIMILARITY).

```
SQ  SEQUENCE   816 AA;  87051 MW;  D49BA5DB423D0777 CRC64;

    Query Match          4.4%;  Score 162; DB 1; Length 816;
    Best Local Similarity  23.0%;  Pred. No. 0.13;
    Matches   85;  Conservative   32;  Mismatches  130;  Indels  122;  Gaps   17;

Qy    397 PPLPHVVVKQRGDSEK-------TDQHKMESSA---ENVVGRLSNQGHHQQSNYMPFANNPP 448
              :  : : :     : ||  : : |:|       : :   | | | :    :
Db     14 PKKREIPATSRSSEEKAPTLPSDNHRVEGTAWLPGNPGGRGHGGGRH----------GPA  63

Qy    449 ASPAPNGY--------------CFPPQPPSGNHQQWLIPVMSPSEGLIYKPHPG------- 489
              | | |||               | |||| | |  | : |  ::  | | ||
Db     64 GTSVELGLQQGIGLHKALSTGLDYSPPSAPRS-------VPVATTLPAAYATPQPGTPVS 116

Qy    490 ---MAHTGH-----------YGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYPP-------- 529
                 ||| |           | |   | | | ||  : || || | || ||
Db    117 PVQYAHLPHTFQFIGSSQYSGTYASFIPSQLIPPTANPVTSAVASAAGATTPSQRSQLEA 176

Qy    530 YGMMPTIMNP------YCSSQQQQQQPNEQMNQFGHPGNLQNTQQQQQR---------S 574
              : |||| ||      :||||||||||||||||||||||||||||||||
Db    177 YSTLLANMGSLSQTPGHKAEQQQQQQQQHQQQQQQQQQQQQQQQQQHLSRAPGLIT 236

Qy    575 DNEPAPQQQQQPTKSYPRARKSRQ-GSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNA 632
              : :   ||| |  :| : |  | | | |  | |  : || |: || :  : :
Db    237 PGSPPPAQQNQ----YVHISSPQNTGRTASPPAIP----------------VHLH 272

Qy    633 PEQTMT--TTTTTRTTTVTQTTRDGGGVTRVIKVVPHNAKLASENAARIFQSIQ------ 684
              :  :|  ||||  |  || | | | | | : | : : |  |: : :| | |
Db    273 PHQTMIPHTLTLGPPSQVVMQYADSGS----HFVPREATYKAE-SSRLQQAIQAKEVLN 326

Qy    685 ---EERKRY 690
              : : | | |
Db    327 GEMEKSRRY 335


RESULT 18
ABF1_MOUSE
ID   ABF1_MOUSE         STANDARD;      PRT;   3726 AA.
AC   Q61329;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2003 (Rel. 42, Last annotation update)
DE   Alpha-fetoprotein enhancer binding protein (AT motif-binding factor)
DE   (AT-binding transcription factor 1).
GN   ATBF1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/MK X ICR; TISSUE=Brain;
RX   MEDLINE=96194902; PubMed=8654949;
RA   Ido A., Miura Y., Watanabe M., Sakai M., Inoue Y., Miki T.,
RA   Hashimoto T., Morinaga T., Nishi S., Tamaoki T.;
RT   "Cloning of the cDNA encoding the mouse ATBF1 transcription factor.";
RL   EMBO J. 16:227-231(1996).
RN   [2]
RP   INTERACTION WITH FNBP3.
RX   MEDLINE=97315177; PubMed=9171351;
RA   Bedford M.T., Chan D.C., Leder P.;
RT   "FBP WW domains and the Abl SH3 domain bind to a specific class of
RT   proline-rich ligands.";
RL   EMBO J. 16:2376-2383(1997).
CC   -!- FUNCTION: Transcriptional activator that binds to the AT-rich core
CC       sequence of the enhancer element of the AFP gene.
CC   -!- SUBUNIT: Interacts with FNBP3.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- SIMILARITY: Contains 4 homeobox domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute.  There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; D26046; BAA05046.1; -.
DR   HSSP; P20263; 1OCP.
DR   MGD; MGI:99948; Atbf1.
DR   TRANSFAC; T03881; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003700; F:transcription factor activity; TAS.
DR   GO; GO:0030182; P:neuron differentiation; TAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001356; Homeobox.
DR   InterPro; IPR007087; Znf_C2H2.
DR   Pfam; PF00046; homeobox; 4.
DR   Pfam; PF00096; zf-C2H2; 18.
DR   ProDom; PD000010; Homeobox; 4.
DR   PROSITE; PS00027; HOMEOBOX_1; 2.
DR   PROSITE; PS00028; ZINC_FINGER_C2H2_1; 15.
DR   PROSITE; PS50071; HOMEOBOX_2; 4.
DR   PROSITE; PS50157; ZINC_FINGER_C2H2_2; 9.
KW   Transcription regulation; Activator; Zinc-finger; Metal-binding;
KW   DNA-binding; Homeobox; Nuclear protein; Repeat.
FT   ZN_FING       79    103       C2H2-TYPE.
FT   ZN_FING      282    305       C2H2-TYPE.
FT   ZN_FING      641    664       C2H2-TYPE.
FT   ZN_FING      672    695       C2H2-TYPE.
FT   ZN_FING      727    751       C2H2-TYPE.
FT   ZN_FING      805    829       C2H2-TYPE.
FT   ZN_FING      946    969       C2H2-TYPE  (DEGENERATE).
FT   ZN_FING      985   1009       C2H2-TYPE  (ATYPICAL).
FT   ZN_FING     1041   1065       C2H2-TYPE  (ATYPICAL).
FT   ZN_FING     1089   1113       C2H2-TYPE  (ATYPICAL).
FT   ZN_FING     1233   1256       C2H2-TYPE  (ATYPICAL).
FT   ZN_FING     1262   1285       C2H2-TYPE.
FT   ZN_FING     1370   1395       C2H2-TYPE.
FT   ZN_FING     1411   1433       C2H2-TYPE.
FT   ZN_FING     1439   1462       C2H2-TYPE.
FT   ZN_FING     1555   1579       C2H2-TYPE.
FT   ZN_FING     1606   1630       C2H2-TYPE.
FT   ZN_FING     1990   2013       C2H2-TYPE.
FT   DNA_BIND    2152   2211       HOMEOBOX 1.
FT   DNA_BIND    2249   2308       HOMEOBOX 2.
FT   ZN_FING     2335   2358       C2H2-TYPE  (ATYPICAL).
FT   ZN_FING     2539   2561       C2H2-TYPE.
FT   DNA_BIND    2650   2709       HOMEOBOX 3.
FT   ZN_FING     2720   2743       C2H2-TYPE.
FT   DNA_BIND    2952   3011       HOMEOBOX 4.
FT   ZN_FING     3032   3056       C2H2-TYPE.
FT   ZN_FING     3552   3576       C2H2-TYPE.
FT   DOMAIN       461    491       POLY-GLU.
FT   DOMAIN       771    785       POLY-ALA.
FT   DOMAIN      1314   1317       POLY-ALA.
FT   DOMAIN      1734   1748       POLY-GLN.
FT   DOMAIN      1794   1799       POLY-GLN.
FT   DOMAIN      1856   1863       POLY-GLN.
FT   DOMAIN      2044   2059       POLY-PRO.
FT   DOMAIN      2405   2408       POLY-ALA.
FT   DOMAIN      3216   3220       POLY-PRO.
FT   DOMAIN      3380   3409       POLY-GLN.
FT   DOMAIN      3412   3420       POLY-GLY.
FT   DOMAIN      3534   3550       POLY-PRO.
FT   DOMAIN      3620   3623       POLY-SER.
FT   DOMAIN      3659   3662       POLY-SER.
SQ   SEQUENCE   3726 AA;  406567 MW;  915ACBE588A72C98 CRC64;

    Query Match          4.4%;  Score 162; DB 1; Length 3726;
    Best Local Similarity  20.7%;  Pred. No. 0.81;
    Matches  147;  Conservative   66;  Mismatches  209;  Indels  288;  Gaps   36;

Qy     43 SIPSQRFGDHGTMNSRSNNTSTLVHP-GPSSQPCGVERNLSVQHLDSSAANQATEKFVSQ 101
              ::  | |             ::   |::| |   :  ||:    :::: ::  :
```

```
Db   2908  SFYSKEYDNEGTVD--YSETSSLADPCSPSPGASG------SAGKSGDGDRPGQKRFRTQ 2960
                 :| ||: |||.   ||:||                    |||
Qy    102  MSFME-NVRSSAQHDQRKMVREE------EDFAVP------VYINSRRSQSHGRTKSGIEKE  150
           |: ||.||| | .| |||:           :||:                  |:|||
Db   2961  MTNLQLKVLKSCFNDYRTPTMLECEVLGNDIGLPKRVVQVWFQNARAK---EKKSKLSMA  3017

Qy    151  KHTPMVAPSSHHSIRFQEVNQT--GSKQNVCLATCSKPEVRDQVKANARSGGFVISLDV   207
                                      |:||
Db   3018  KHF-------------GINQTSYEGPKTECTL--CG------IKYSAR---------L  3045

Qy    208  SVTEIDLEKSASSHDRVNDYNASLRQESNRLYRDGGKTRLKDTDNGAESHLATENHSQ   267
                                                          ::|||
Db   3046  SVRDHIFSQQHIS-----------------KVKDTIG----------------------  3065

Qy    268  EGHGSPEDIDNDREYSKSRACASLQQINEEASDVSDSMVDSISSIDVSPDDVVGILGQ   327
                                                         |||
Db   3066  -----SQLDKEKEY--------------------------FDPATVRQLMAQ  3086

Qy    328  KRFWRARKA------IANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEIS------FL  375
                .: |||                                     :: |:
Db   3087  QELDRIKKANEVLGLAAQQQGMFDNAPLQALNLPTTYPALQGIPVLLPGLNRPSLPGFT  3146

Qy    376  GKVSAKSYPVKKL-LPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQG   433
                |          :|| ||:|                         :
Db   3147  PANTALITSPKFNLMGLPSTTVPSPGLPTSGLPNKPSS-----------ASLSSPT  3190

Qy    434  HHQQSNYMPFANNPPASP-APNGYCFPP--QPPPSGNHQQWLIPVMSPSE-------  480
                |           |.                 |||
Db   3191  PAQAT--MAMAPQPPPQPQQPQ-----PPVQQPPPPAAQQIPAPQITPQQQRKDGEKG  3244

Qy    481  --------------------GLIYKPHPGM----------AHTGHYG  497
                                    |:||||
Db   3245  KEKEKAHKGKGEPLPVPKKKGEAPPAGTGTISAPLPAMEYAVDPAQLQAIQAALTSD--  3302

Qy    498  GYYGHYMPTPMVMPQYHPGM-GFPP------PG---NGYFPP-YGM-----MPTIMNP  539
                :|| |||||                ||:|| |:  :||||   |||
Db   3303  -------PTAILLTSQFLPYFVPGFSPYYAPQIPGALQSGYLQPMYGMEGLFPYSPALSRP  3355

Qy    540  ---------YCSS------QQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPA  579
                                 ||||
Db   3356  LMGLSPGSLLQQYQQYQQSLQEAIQQQQQQQQQQQQRQLQQQQQQQQQKVQQ-Q  3414

Qy    580  PQQQQQPTKSYPRARKSRQGTGSSPSGPQGISGSKSFRPFAAVDEDSNI  629
                ||||||
Db   3415  QQQQQQP--------KASQTPVPQGAASPDKDPAKESPKP---EEQKNV  3452
```

RESULT 19
TF20_HUMAN

ID   TF20_HUMAN   STANDARD;   PRT;   1960 AA.
AC   Q9UGU0; O14528; Q13078; Q9H4M0;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Transcription factor 20 (Stromelysin 1 PDGF-responsive element-binding
DE   protein) (SPRE-binding protein) (Nuclear factor SPBP) (AR1).
GN   TCF20 OR SPBP OR KIAA0292.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), CHARACTERIZATION, AND
RP   ALTERNATIVE SPLICING.
RX   MEDLINE=20568288; PubMed=10995766;
RA   Rekdal C., Sjoettem E., Johansen T.;
RT   "The nuclear factor SPBP contains different functional domains and
RT   stimulates the activity of various transcriptional activators.";
RL   J. Biol. Chem. 275:40288-40300(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20057165; PubMed=10591208;
RA   Dunham I., Hunt A.R., Collins J.E., Bruskiewich R., Beare D.M.,
RA   Clamp M., Smink L.J., Ainscough R., Almeida J.P., Babbage A.K.,
RA   Bagguley C., Bailey J., Barlow K.F., Bates K.N., Beasley J.,
RA   Bird C.P., Blakey S.E., Bridgeman A.M., Buck D., Burgess J.,
RA   Burrill W.D., Burton J., Carder C., Carter N.P., Chen Y., Clark G.,
RA   Clegg S.M., Cobley V.E., Cole C.G., Collier R.E., Connor R.,
RA   Conroy D., Corby N.R., Coville G.J., Cox A.V., Davis J., Dawson E.,
RA   Dhami P.D., Dockree C., Dodsworth S.J., Durbin R.M., Ellington A.G.,
RA   Evans K.L., Fey J.M., Fleming K., French L., Garner A.A.,
RA   Gilbert J.G.R., Goward M.E., Grafham D.V., Griffiths M.N.D., Hall C.,
RA   Hall R.E., Hall-Tamlyn G., Heathcott R.W., Ho S., Holmes S.,
RA   Hunt S.E., Jones M.C., Kershaw J., Kimberley A.M., King A.,
RA   Laird G.K., Langford C.F., Leversha M.A., Lloyd C., Lloyd D.M.,
RA   Martyn I.D., Mashreghi-Mohammadi M., Matthews L.H., Mccann O.T.,
RA   Mcclay J., Mclaren S., McMurray A.A., Milne S.A., Mortimore B.J.,
RA   Odell C.N., Pavitt R., Pearce A.V., Pearson D., Phillimore B.J.C.T.,
RA   Phillips S.H., Plumb R.W., Ramsay H., Ramsey Y., Rogers L., Ross M.T.,
RA   Scott C.E., Sehra H.K., Skuce C.D., Smalley S.J., Smith M.L.,
RA   Soderlund C., Spragon L., Steward C.A., Sulston J.E., Swann R.M.,
RA   Vaudin M., Wall M., Wallis J.M., Whiteley M.N., Willey D.L.,
RA   Williams L., Williams S.A., Williamson H., Wilmer T.E., Wilming L.,
RA   Wright C.L., Hubbard T., Bentley D.R., Beck S., Rogers J., Shimizu N.,
RA   Minoshima S., Kawasaki K., Sasaki T., Asakawa S., Kudoh J.,
RA   Shintani A., Shibuya K., Yoshizaki Y., Aoki N., Mitsuyama S.,
RA   Roe B.A., Chen F., Fu Y., Hu P., Hua A., Kenton S., Lai H., Lao H.I.,
RA   Dorman A., Fang F., Fu Y., Hua A., Kenton S., Lai H., Lao H.I.,
RA   Lewis J., Lewis S., Lin S.-P., Loh P., Malaj E., Nguyen T., Pan H.,
RA   Phan S., Qi S., Qian Y., Ray L., Ren Q., Shauli I., Sloan D., Song L.,
RA   Wang Q., Wang Y., Wang Z., White J., Willingham D., Wu H., Yao Z.,
RA   Zhan M., Zhang G., Chissoe S., Murray J., Miller N., Minx P.,
RA   Fulton R., Johnson D., Bemis G., Bentley D., Bradshaw H., Bourne S.,
RA   Cordes M., Du Z., Fulton L., Goela D., Graves T., Hawkins J.,
RA   Hinds K., Kemp K., Latreille P., Layman D., Ozersky P., Rohlfing T.,
RA   Scheet P., Walker C., Wamsley A., Wohldmann P., Pepin K., Nelson J.,
RA   Korf I., Bedell J.A., Hillier L.W., Mardis E., Waterston R.,
RA   Wilson R., Emanuel B.S., Shaikh T., Kurahashi H., Saitta S.,
RA   Budarf M.L., McDermid H.E., Johnson A., Wong A.C.C., Morrow B.E.,
RA   Edelman L., Kim U.J., Shizuya H., Simon M.I., Dumanski J.P.,
RA   Peyrard M., Kedra D., Seroussi E., Fransson I., Tapia I., Bruder C.E.,
RA   O'Brien K.P., Wilkinson P., Bodenteich A., Hartman K., Hu X.,
RA   Khan A.S., Lane L., Tilahun Y., Wright H.;
RT   "The DNA sequence of human chromosome 22.";
RL   Nature 402:489-495(1999).
RN   [3]
RP   SEQUENCE OF 245-1960 FROM N.A. (ISOFORM 1).
RC   TISSUE=Brain;
RX   MEDLINE=97323006; PubMed=9179496;
RA   Ohara O., Nagase T., Ishikawa K.-I., Nakajima D., Ohira M.,
RA   Seki N., Nomura N.;
RT   "Construction and characterization of human brain cDNA libraries
RT   suitable for analysis of cDNA clones encoding relatively large
RT   proteins.";
RL   DNA Res. 4:53-59(1997).
RN   [4]
RP   SEQUENCE OF 1083-1938 FROM N.A. (ISOFORM 2).
RX   MEDLINE=98399843; PubMed=9730594;
RA   Rajadhyaksha A., Riviere M., Van Vooren P., Szpirer J., Szpirer C.,
RA   Babin J., Bina M.;
RT   "Assignment of AR1, transcription factor 20 (TCF20), to human
RT   chromosome 22q13.3 with somatic cell hybrids and in situ
RT   hybridization.";
RL   Cytogenet. Cell Genet. 81:176-177(1998).
CC   -!- FUNCTION: Transcriptional activator that binds to the regulatory
CC   region of MMP3 and thereby controls stromelysin expression. It
CC   stimulates the activity of various transcriptional activators such
CC   as JUN, SP1, PAX6 and ETS1, suggesting a function as a
CC   coactivator.
CC   -!- SUBUNIT: Homodimer (Probable). Interacts with RNF4 and JUN (By
CC   similarity).
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC   Event=Alternative splicing; Named isoforms=2;
CC   Name=1;
CC   IsoId=Q9UGU0-1; Sequence=Displayed;

```
CC        Name=2;
CC        IsoId=Q9UGU0-2; Sequence=VSP 003984, VSP 003985;
CC   -!- TISSUE SPECIFICITY: Expressed in most tissues, except in ovary and
CC       prostate. Isoform 1 is exclusively expressed in brain, heart and
CC       testis, and this form predominates in liver and kidney. Isoform 2
CC       predominates in lung.
CC   -!- DOMAIN: The atypical PHD domain functions as a negative modulator
CC       of cofactor binding (By similarity).
CC   -!- SIMILARITY: Contains 1 A.T hook DNA-binding repeat.
CC   -!- SIMILARITY: Contains 1 PHD-type zinc finger.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to a
CC       frameshift in position 1932.
CC   -----------------------------------------------------------------------
CC       This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------------
DR   EMBL; AY007595; AAG28930.1; -.
DR   EMBL; AL031346; CAB42440.1; ALT_INIT.
DR   EMBL; AL021878; -; NOT_ANNOTATED_CDS.
DR   EMBL; AB006630; BAA22961.1; -.
DR   EMBL; U19345; AAC36392.1; ALT_FRAME.
DR   Genew; HGNC:11631; TCF20.
DR   MIM; 603107; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003677; F:DNA binding; NAS.
DR   GO; GO:0003713; F:transcription co-activator activity; NAS.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   InterPro; IPR001965; Znf_PHD.
DR   SMART; SM00249; PHD; 1.
KW   Transcription regulation; Activator; DNA-binding; Zinc-finger;
KW   Nuclear protein; Alternative splicing.
FT   DOMAIN        1170       1191       LEUCINE-ZIPPER.
FT   DOMAIN        1254       1268       NUCLEAR LOCALIZATION SIGNAL
FT                                       (BY SIMILARITY).
FT   DNA_BIND      1537       1551       A.T HOOK.
FT   DOMAIN        1576       1600       NUCLEAR LOCALIZATION SIGNAL
FT                                       (BY SIMILARITY).
FT   DOMAIN        1785       1792       NUCLEAR LOCALIZATION SIGNAL
FT                                       (BY SIMILARITY).
FT   ZN_FING       1884       1935       PHD-TYPE (ATYPICAL).
FT   DOMAIN          62         70       POLY-ALA.
FT   DOMAIN         174        182       POLY-GLN.
FT   DOMAIN         203        262       SER-RICH.
FT   DOMAIN         310        322       POLY-GLN.
FT   DOMAIN        1556       1564       POLY-PRO.
FT   VARSPLIC      1934       1938       PPLPC -> VRLWR (in isoform 2).
FT                                       /FTId=VSP_003984.
FT   VARSPLIC      1939       1960       Missing (in isoform 2).
FT                                       /FTId=VSP_003985.
FT   CONFLICT       122        122       R -> Q (IN REF. 2).
FT   CONFLICT       200        200       K -> Q (IN REF. 2).
SQ   SEQUENCE   1960 AA;  211798 MW;  FB459146C05B6D2A CRC64;
```

```
Query Match             4.4%;  Score 161;  DB 1;  Length 1960;
Best Local Similarity  19.1%;  Pred. No. 0.43;
Matches  147;  Conservative  90;  Mismatches  261;  Indels  270;  Gaps  32;

Qy    17 RLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSRSNNTSTLVHPGPSSQP-- 74
                                                            |:|:
Db  1015 KLKMSPGRSRGPGGDPHH-----------MNPHMTFSERANRSSLHTPFSPNSETLA 1060

Qy    75 ---------CGVERNLSVQHLDSSAANQATEKFVSQMS-FMENVRSSAQHDQ- 116
                  :|                  |:          ||      |:
Db  1061 SAYHANTRAHAYGDPNAGLNSQL--HYKRQMYQQOPEYKDWSSGAQGVIAAAQHRQE 1117

Qy   117 --RKMVREEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPS---------------- 159
            ||                  :||   : ||   |  ::: |
Db  1118 GPRK------------SPRQQOFLDRVRSPLKNDKDGMMYGPPVGTYHDFSAQEAGRC 1163
```

```
Qy   160 --------------------SHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQ--------- 191
                             |::|||:    |:|::|
Db  1164 LMSSDGLPNKGMELKHGSQKLQESCWDLSRQTSPAKSSGPPGMSSQKRYGPPHETDGHGL 1223

Qy   192 VKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASL--RQESRNLYRDGK 246
             |:|      |:|    :        ::|:     |::
Db  1224 AEATQSSKPGSVMLRLPGQEDHSSQNPLIMRRRVRSFISPIFSKRQSQDVKNSSTEDKGR 1283

Qy   247 TRLKDTDNGAE------SHLATENHSQEGHGSPE-DIDNDREYSKSRACASLQQINEEAS 299
             | :|||             |:|:|      |: |:|::        |:    ||  :
Db  1284 L-LHSSKEGADKAFNSYAHL--SHSQDIKSIPKRDSKDLPSDSRNCPAV---------- 1331

Qy   300 DDVSDSDSMVDSISIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHRLIKVQK 359
             |  :|                ::|  |:|           |:    |  |:|
Db  1332 -TLTSPAKTKILP----------PRKGRGLKLEAIVQKITSP--NIRRSASSNS 1372

Qy   360 LIAASPDLLLDEISFL--------GKVSAKSYPVKK-------LLPS--EFLVKPPLPH 401
             :|  |  |:|::       :||   |      |            ||    ||:|||
Db  1373 AEAGGDTVTLDDILSLKSGPPEGGSVAVQDADIEKRKGEVASDLVSPANQELHVEKPLPR 1432

Qy   402 VVVKQRGDSEKTDQHKMESSAENVV------GRL----SNQHHQQS-----NY 440
            ||:|:|  |:  | |:::|                 |:|     ::
Db  1433 SSEEWRGSVD--DKVKTETHARTVTAGKEPPGAMTSTTSQKPGSNQRPDGSLGGTAPLI 1490

Qy   441 MPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYKFPHPGMAHTGHYGGYY 500
                   ||:|      |  :|| |  :| |::|:         |  |
Db  1491 FPDSKNVP----PVGILAPEANPKAEEKEND`VTISPKQE-- 1526

Qy   501 GHYMPTPMVMPQYHPGMGFPPPPGNGYFP-----------PYGMMPTIMNPYCSSQQQQQQPNE 553
             |         ||                       ||::      |:  |
Db  1527 ----------GFPP--KGYFPSGKKGRPIGSV--------NKQKKQQQP-- 1556

Qy   554 QMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTK----SYPRARKSRQGSTGSSPSGP 608
                                              |:|||:
Db  1557 ------PPPPQPPQIPEGSADGEPKPKQRRERRKPGAQPRKKRTKQAVPIVEFQEP 1610

Qy   609 Q--------------------GISGSKSFRPFAAVDEDSNIN------NAPEQTMT 638
                                  :||:|::                   |||:|
Db  1611 EIKLKYATQPLDKTDAKNKSFYPYIHVVNKCELGAVCTIINAEEEQT 1658
```

```
RESULT 20
Y553 HUMAN
ID   Y553_HUMAN              STANDARD;      PRT;   1089 AA.
AC   Q9UKJ3; O60300;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein KIAA0553.
GN   KIAA0553.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=98290545; PubMed=9628581;
RA   Nagase T., Ishikawa K.-I., Miyajima N., Tanaka A., Kotani H.,
RA   Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. IX.
RT   The complete sequences of 100 new cDNA clones from brain which can
RT   code for large proteins in vitro.";
RL   DNA Res. 5:31-39(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=99408744; PubMed=10477733;
RA   Thornton M.A., Poncz M., Korostishevsky M., Yakobson E., Usher S.,
RA   Seligsohn U., Peretz H.;
RT   "The human platelet alphaIIb gene is not closely linked to its
RT   integrin partner beta3.";
RL   Blood 94:2039-2047(1999).
```

```
CC    -----------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC    -----------------------------------------------------------------
DR    EMBL; AB011125; BAA25479.1; ALT_INIT.
DR    EMBL; AF160252; AAF03681.1; -.
KW    Hypothetical protein.
FT    DOMAIN      263    266       POLY-LYS.
FT    DOMAIN      332    336       POLY-ARG.
FT    DOMAIN      359    362       POLY-GLY.
FT    DOMAIN      425    430       POLY-GLU.
FT    DOMAIN      456    461       POLY-SER.
FT    DOMAIN      971    977       POLY-ALA.
FT    DOMAIN     1059   1065       POLY-ALA.
FT    CONFLICT    207    207       K -> E (IN REF. 1).
FT    CONFLICT    267    267       E -> K (IN REF. 1).
FT    CONFLICT    278    278       P -> H (IN REF. 1).
FT    CONFLICT    846    846       G -> S (IN REF. 1).
SQ    SEQUENCE   1089 AA;  117999 MW;  4EF687F9D81A16A5 CRC64;

Query Match             4.3%;  Score 158;  DB 1;  Length 1089;
Best Local Similarity  18.7%;  Pred. No. 0.31;
Matches  146;  Conservative  115;  Mismatches  309;  Indels  210;  Gaps   36;

Qy     2 KRGKDEEKILEPMFPRLHVNDADKG-------GPRAPPRNKMALY--EQLSIPSQRFG  50
         ||  |    ::| |:    : |::          |:::||                  :
Db   300 KRKRKNKSSAPA------DSERGPKPEPPGSGSPAPPRRRRAQDDSQRRSLPAEE-G 351

Qy    51 DHGTMNSRSNNTSTLVHPG---------PSS--QPCGVERNLSVQH----------LDSSA  90
         ||                          |:|   :| :::                 ::
Db   352 SSGKKDEGGGSSSQDHGGRHKGELPPSSCQRRAGTKRSRSSHRSQPSSGDEDSDDAS 411

Qy    91 ANQATEKFVSQMSFME------------NVRSSAQHDQRKMVR---------------- 121
            :::|                     ||| ||
Db   412 SHRLHQKSPSQYSEEEEEEDSGSEHSRSRSGRRHSSHRSSRRSYSSSDASSDQSCYS 471

Qy   122 EEEDFAVPVYIN-SRRSQSHGRTKSGIEKEKTPMVAPSSHHSIRFQEVNQTGSKQNVCL 180
          |: |    :     ::   |::    :: | : :::   ::    : | |::||::|
Db   472 RQRSYSDDSYSDYSDRSRRHSKRSHDSDDSDY---ASSKHRSKRHKYSSSDDDYSLSC- 526

Qy   181 ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSH--DRVNDYNASLRQESRN 238
         |  | |   : :  :: ::|| |   :  :    : : ::|     : :|| ||:| :|
Db   527 -SQSRSRSRHTRERSRSRGRSRSSSCSRSRSKRRSRSTTAHSWQRSRSYSRDRSRSTRS 585

Qy   239 RLYRDGGKTRLKDTNGAESHLATEN-------HSQ---------EGHGSPEDIDNDRE 281
         || |   |     :| :   |  |:         :|         :|  |  :|  |:
Db   586 PSQRSGSRKRSWGHESPPERHSGRRDFIRSKIYRSQSPHYFRSGRGEGPGKKDDGRGD-- 643

Qy   282 YSKSRACASLQQINEEASDDVSDDSMVDS--ISSIDVSPDDVVGILGQKRFWRARKAIAN 339
           |::|| :|   :: ::  |                :||| : ::||   :|  |:|
Db   644 --------DSKATGPPSQNSNIGTGRGSEGDCSPED-----------KNSVTA 677

Qy   340 QQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKPPL 399
         |||  |  | ::||| |   ||:      ||: : |:       |  :: || || |
Db   678 KLLLEKIQSRKVERKPSVSEEVQATPNKAGPKL----KDPPQGYFGPKLPPS-LGNKPVL 732

Qy   400 PHVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPPANNPP-ASPAPNGYCF 458
          |: :||||  | :    |:::|      |  :    : :| :  ||    : |::| |
Db   733 P--LIGKLPATRKPNKKCEESGLE---------RGEEQEQS--ETEEGPPGSSDALFGHOF 780

Qy   459 PPQ------PPP-----SGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGH 502
          ||        |       ::| ||      ::|| |    | |  :||  :  |
Db   781 PSEETTGPLLDPPPEESKSGEATADHP--VAPLGTPAHSDCYPGDPTISH-------N 829

Qy   503 YMPTPM---VMPQXHPGMGFPP------PGNGYFPPYGMGPTIMNPYCSS----QQQQ 547
         |: |      ||  ||  |   |      ||   |        : : : ::
Db   830 YLPDPSDGDTLESLDSGSQPGFVESSLLPIAPDLLEHFPSYA--PPSGDFSIESTDGAEDA 887

Qy   548 QQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSG 607
         ||                    :  :   |                :  :|
Db   888 SLAPLESQPITFTPEEMEKYSKLQQAAQQHIQQQLLAKQVKAFP------ASAALAPAT 940

Qy   608 PQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRDGGGVTRVIKVVPH 667
          |: ||  ||  |                :    ::  | | |: | |  | |::|  ||
Db   941 P------ALQPIHI-------QQPATASATSITTVQHAILQ--HHAAAAAAIGIHPH 983
```

```
RESULT 21
CBP1_CAEEL
ID   CBP1_CAEEL              STANDARD;      PRT;  2056 AA.
AC   P34545;
DT   01-FEB-1994 (Rel. 28, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein cbp-1.
GN   CBP-1 OR R10E11.1.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,
RA   Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,
RA   Johnston L., Jones M., Kershaw J., Kirsten J., Laister N.,
RA   Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,
RA   Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,
RA   Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,
RA   Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
RN   [2]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RA   Durbin R.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=b;
CC         IsoId=P34545-1; Sequence=Displayed;
CC       Name=a;
CC         IsoId=P34545-2; Sequence=VSP_000557;
CC         Note=No experimental confirmation available;
CC   -!- SIMILARITY: Contains 1 bromodomain.
CC   -!- SIMILARITY: Contains 1 ZZ-type zinc finger.
CC   -!- SIMILARITY: Contains 1 TAZ-type zinc fingers.
CC   -----------------------------------------------------------------
CC    This SWISS-PROT entry is copyright. It is produced through a collaboration
CC    between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC    the European Bioinformatics Institute. There are no restrictions on its
CC    use by non-profit institutions as long as its content is in no way
CC    modified and this statement is not removed. Usage by and for commercial
CC    entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC    or send an email to license@isb-sib.ch).
CC   -----------------------------------------------------------------
DR   EMBL; Z29095; CAA82353.2; -.
DR   EMBL; Z29095; CAD18875.1; -.
DR   PIR; G88563; G88564.
DR   WormPep; R10E11.1a; CE28069.
DR   WormPep; R10E11.1b; CE21117.
DR   InterPro; IPR001487; Bromodomain.
DR   InterPro; IPR003101; KIX.
DR   InterPro; IPR000197; TAZ_finger.
DR   InterPro; IPR001965; Znf_PHD.
DR   InterPro; IPR000433; Znf_ZZ.
```

```
DR   Pfam; PF00439; bromodomain; 1.
DR   Pfam; PF02172; KIX; 1.
DR   Pfam; PF02135; zf-TAZ; 2.
DR   Pfam; PF00569; ZZ; 1.
DR   PRINTS; PR00503; BROMODOMAIN.
DR   SMART; SM00297; BROMO; 1.
DR   SMART; SM00551; ZnF_TAZ; 2.
DR   SMART; SM00291; ZnF_ZZ; 1.
DR   PROSITE; PS00633; BROMODOMAIN_1; 1.
DR   PROSITE; PS50014; BROMODOMAIN_2; 1.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50134; ZF_TAZ; 2.
DR   PROSITE; PS01357; ZF_ZZ_1; 1.
DR   PROSITE; PS50135; ZF_ZZ_2; 1.
KW   Bromodomain; Metal-binding; Zinc; Zinc-finger; Alternative splicing;
KW   Repeat.
FT   ZN_FING       399    505       TAZ-TYPE 1.
FT   DOMAIN        881    953       BROMODOMAIN.
FT   ZN_FING      1493   1534       ZZ-TYPE.
FT   ZN_FING      1550   1631       TAZ-TYPE 2.
FT   DOMAIN       1687   2008       GLY/GLN-RICH.
FT   VARSPLIC      467    478       SDTTQTTKKCSV -> F (in isoform a).
FT                                  /FTId=VSP_000557.
SQ   SEQUENCE   2056 AA;  227179 MW;  949FF46D8C634F01 CRC64;
```

```
Query Match             4.3%;
Best Local Similarity  24.9%;
  Matches   75;  Conservative   29;  Mismatches  117;  Indels   80;  Gaps   16;

Qy    439 NYMPFANNPPA------SPAPNGYCFPPQPPPSGNHQQWLIPVMSPSE------------- 480
Db    136 NGTPNMMSPPSMGRVPGPSPGG---PQPPGPGQ----PQMRPGQPGMFQGDQQQQMM 185
Qy    481 -GLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGM--GFPPPG-----------NGYPP 528
Db    186 MGAQGQQPFGMMHRYPYA--QGGPPPGAQGMPQGYPGVSRGGPTPGQPMGRGAMMNGAMP 243
Qy    529 PYGMMPT----------IMNPYCSSQQ----QQQQQPNEQMNQFGHPGNLQNTQQ-Q 570
Db    244 RSGPMPTQGRPGIPPNQQAMMQPMMTDRQFMQHGQYGQQRPEFMQQYGRPGGYPMMHQGM 303
Qy    571 QQRSDNEP--APQQQQQPTKSYPRARKSRQGSTGSSPS---GPQGISGSKSFRPFAAVDE 625
Db    304 MMDSNGQPIRGPNQMMMMSNGHP------GMSHGPPNGQPGPQAAAAQHAAQQQAAAQA 356
Qy    626 DSNINNAPEQTMTTTTTTTTRTTVTQTRTDGGGVTRVIKVVPHNAKLASENAARIFQSIQE 685
Db    357 QAQA--AAQQQQQQREQEAAAAAQ--RNGAGRA----TTPGSSMLATHQDPEKRKLIQQ 408
Qy    686 E 686
          :
Db    409 Q 409
```

```
RESULT 22
PROS_DROME
ID   PROS_DROME     STANDARD;      PRT;   1403 AA.
AC   P29617; Q95SP0; Q9U6A2; Q9VGP8;
DT   01-APR-1993 (Rel. 25, Created)
DT   01-FEB-1996 (Rel. 33, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein prospero.
GN   Name OR CG17228.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92069760; PubMed=1720353;
RA   Vaessin H., Grell E., Wolff E., Bier E., Jan L.Y., Jan Y.N.;
RT   "Prospero is expressed in neuronal precursors and encodes a nuclear
```

```
RT   protein that is involved in the control of axonal outgrowth in
RT   Drosophila.";
RL   Cell 67:941-953(1991).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=92171948; PubMed=1540176;
RA   Matsuzaki F., Koizumi K., Hama C., Yoshioka T., Nabeshima Y.;
RT   "Cloning of the Drosophila prospero gene and its expression in
RT   ganglion mother cells.";
RL   Biochem. Biophys. Res. Commun. 182:1326-1332(1992).
RN   [3]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93083413; PubMed=1842358;
RA   Chu-Lagraff Q., Wright D.M., McNeil L.K., Doe C.Q.;
RT   "The prospero gene encodes a divergent homeodomain protein that
RT   controls neuronal identity in Drosophila.";
RL   Development Suppl. 2:79-85(1991).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM C).
RX   MEDLINE=20503846; PubMed=11051550;
RA   Xu C., Kauffmann R.C., Zhang J., Kladny S., Carthew R.W.;
RT   "Overlapping activators and repressors delimit transcriptional
RT   response to receptor tyrosine kinase signals in the Drosophila eye.";
RL   Cell 103:87-97(2000).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Celniker S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov P., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [6]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
```

```
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [7]
RP   SEQUENCE FROM N.A. (ISOFORM A).
RC   STRAIN=Berkeley; TISSUE=Head;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
RN   [8]
RP   SIMILARITY TO C.ELEGANS CEH-26.
RX   MEDLINE=94212446; PubMed=7909177;
RA   Buerglin T.R.;
RT   "A Caenorhabditis elegans prospero homologue defines a novel domain.";
RL   Trends Biochem. Sci. 19:70-71(1994).
CC   -!- FUNCTION: Required for proper neuronal differentiation of most or
CC       all neurons and their precursors in central and peripheral nervous
CC       systems, axonal outgrowth and pathfinding. Not required for the
CC       specification of neuronal identity. May regulate transcription by
CC       binding to DNA.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=C; Synonyms=L;
CC         IsoId=P29617-1; Sequence=Displayed;
CC       Name=A;
CC         IsoId=P29617-2; Sequence=VSP_002307, VSP_002308;
CC       Name=B;
CC         IsoId=P29617-3; Sequence=VSP_002307, VSP_002307, VSP_002309;
CC       Name=S; Synonyms=D;
CC         IsoId=P29617-4; Sequence=VSP_002308;
CC   -!- TISSUE SPECIFICITY: Neuronal precursors. Expressed in the
CC       developing CNS, lens-secreting cone cells of the eye, and midgut.
CC   -!- DEVELOPMENTAL STAGE: Expressed in neuronal precursors early during
CC       formation.
CC   -!- SIMILARITY: Belongs to the Prospero homeobox family.
CC   -!- SIMILARITY: Contains 1 homeobox domain.
CC   -!- CAUTION: Ref.7 sequence differs from that shown due to a
CC       frameshift in position 1122.
```

```
DR   EMBL; M81389; AAA28841.1; -.
DR   EMBL; D10609; BAA01464.1; -.
DR   EMBL; Z11743; CAA77802.1; -.
DR   EMBL; AF190403; AAF05703.1; -.
DR   EMBL; AE003691; AAF54628.2; -.
DR   EMBL; AE003691; AAN13500.2; -.
DR   EMBL; AE003691; AAN13501.2; -.
DR   EMBL; AY060680; AAL28228.1; ALT_FRAME.
DR   PIR; S24548; S24548.
DR   PDB; 1MIJ; 04-DEC-02.
DR   FlyBase; FBgn0004595; pros.
DR   GO; GO:0045179; C:apical cortex; IDA.
DR   GO; GO:0045180; C:basal cortex; IDA.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0003700; F:transcription factor activity; NAS.
DR   GO; GO:0007409; P:axonogenesis; IMP.
DR   GO; GO:0007417; P:central nervous system development; IMP.
DR   GO; GO:0016358; P:dendrite morphogenesis; IMP.
DR   GO; GO:0007422; P:peripheral nervous system development; IMP.
DR   GO; GO:0045664; P:regulation of neuron differentiation; IMP.
DR   InterPro; IPR007738; Prox1.
DR   Pfam; PF05044; Prox1; 1.
KW   Nuclear protein; Transcription regulation; DNA-binding; Homeobox;
KW   Developmental protein; Alternative splicing; 3D-structure.
FT   DOMAIN         4     12       POLY-ALA.
FT   DOMAIN        28     31       POLY-SER.
FT   DOMAIN        32     35       POLY-ASN.
FT   DOMAIN       188    191       POLY-ALA.
FT   DOMAIN       216    264       GLN-RICH.
FT   DOMAIN       270    286       ASN-RICH.
FT   DOMAIN       318    354       SER-RICH.
FT   DOMAIN       431    437       POLY-ASP.
FT   DOMAIN       505    508       POLY-ALA.
FT   DOMAIN       700   1048       GLN-RICH.
FT   DOMAIN       934    937       POLY-ALA.
FT   DOMAIN       991    998       NUCLEAR LOCALIZATION SIGNAL.
FT   DOMAIN      1074   1082       POLY-PRO.
FT   DOMAIN      1127   1137       HIS-RICH.
FT   DOMAIN      1241   1303       HOMEOBOX (ATYPICAL).
FT   DNA_BIND    1304   1403       PROSPERO-LIKE.
FT   VARSPLIC     916    916       S -> SGNNGSLLLANSQMPSQTTASGSSAQQQQQNAQQQ
FT                                 HQGSQQQQQQNVVQQNVAQQQHMQQQQQQQSHPLLPPNC
FT                                 QQLISAPRLNGSQLSFASPAAAAAAMGLQMHHAAAAAMS
FT                                 AQQQQQSGDPGANTNPNSGPANPTNSLSTLNIPPPHIRP
FT                                 SPT (in isoform A and isoform B).
FT                                 /FTId=VSP_002307.
FT   VARSPLIC    1216   1244       Missing (in isoform A and isoform S).
FT                                 /FTId=VSP_002308.
FT   VARSPLIC    1236   1403       Missing (in isoform B).
FT                                 /FTId=VSP_002309.
FT   CONFLICT      76     98       AKMLNELFGROMKQAQDATSGLP -> GQDAERAVWPPDEA
FT                                 GPGRNEWPA (IN REF. 1 AND 4).
FT   CONFLICT     120    144       IGSLNSTSKLLQQQHNNSIAPANS -> NLALQFHVQVAA
FT                                 AAAITTALLPPIG (IN REF. 1 AND 4).

Query Match                4.3%;  Score 156; DB 1; Length 1403;
Best Local Similarity     17.8%;  Pred. No. 0.53;
Matches  155;  Conservative 138;  Mismatches 336;  Indels 240;  Gaps 36;

Qy     4 GKDEEKILEPMFPR--LHVNDADKGGPRAPPRNKMA----------------LYEQLSIPSQRFG 50
           :|:|| :|||||   :|:||  ||::                          ||:|: :
Db    72 GGSSAKMLNELFGRQMKQAQDATSGLPQSLDNAMLAAAMETATSAELLIGSLNSTSKLLQ 131

Qy    51 DHGTMNS----------------RSNNTSTLVHPGSS----------QPCG------VERNLSVQHL 86
           : | |:                :|:|:|:|:|||               |||      ::|
Db   132 QQHNNNSIAPANSTPMSNGTNASISPGSAHSSSHSHQGVSPKGSRRVSACSDRSLEAAAA 191

Qy    87 DSSAANQATEKFVSQM---SFMENVRSSAQHD--------------QRKMVREEEDFAVP 129
           |  :: :: :: ::                                ::::: :::
Db   192 DVAGGSPPRAASVSSLNGGASSGEQHQSOLQHDLVAHHMLRNILQGKKELMQLDQELRTA 251

Qy   130 VYINSRRSQSHGRTKSGIEKEKHTPMVA-PSSHHSIRFQEVNQTGSKQNVCLATCSKPEV 188
           |::|| | :: :  || :   :: ||   :::||  ||:  : : :|:::::::::::
Db   252 MQQQQQLQEKEQLHSKLNNNNNNNIAATANNNNNTTMESINLIDDSEMADIKIKSEPQT 311

Qy   189 RDQVKA------NARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLY 241
           || |:       ::  :::::::||:  ::::||  ::  :::|
Db   312 APQPQQSPHGGSSHSSRSG---------------SGSGGSHSSMAS-DGSLRRKSSDSLD 353

Qy   242 RDGGKTRLKDTDNGAESHLATENHSQEGHGSP---EDID-----------------NDREYS 283
           |::|:| ::| :|  ::: |::|:|:|:|::    :::
Db   354 SHGAQDDAQDEEDAAPTGQRSESRAPEEPQLPTKKESVDDMLDEVLLGLHSRGSDMDSL 413

Qy   284 KSRACASLQQIN-EEASDDVSDDSMVDS------------------ISSIDVSPD 319
           ::| :|:::|: :::| :|:::|:|
Db   414 ASPSHSDMMLLDKDDVLDEDDDDCVEQKTSGSGCLKKPGMDLKRARVENIVSGMRCSPS 473

Qy   320 DVVGIILGQKRFWRARKAIANQQRVPAVQLF----ELHRLIKVQKLIAASPDLLLDEISF 374
           |:||:||: ||: ::|::::: :: |: ::     ::|:::|:| |::|: ::::|:|:
Db   474 SGLAQAGLQLQVNGCKKRKLYQPQOHAMERYVAAAAGLNFGLNLQSMMLDQEDSESNELES 533
```

```
Qy   375  --LGKVSAKSYPVKKLLPS--EFLVKPPLPHVV----VKQRGDSEKTDQH--  423
                :   ::   :   ::    :  ::         :: :
Db   534  PQIQQKRVEKNALKSQLRSMQEQLAEMQQKYVQLCSRMEQESECQELDQQDVEQEQEPD  593

Qy   424  NVVGRLSNQGHHQQSNYMPFANNPPASPAP---NGYCFPPQPPPSGNHQQWLIPVMSPS  479
                                          :::
Db   594  N----GSSDHIE--------LSPSP'TLTGDGVSPNHKEETGQER----PGSSSP  632

Qy   480  EGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYH--PGMGFPPPGNGYFPP---YGM  532
          ::::
Db   633  SPSPIKPKTSLGESSDSGANMLSQMMSKMMSGKLHNPLVGVHPALPQG-FPPLLQHMGD  691

Qy   533  MPTIMNPY---------------CSSQQQQQQPNEQMNQFGHPGNLQNTQQQQRSD  575
                           ||||||   :      |     |  ||
Db   692  MSHAAAMYQQFFFEQEARMAKEAEAQQQQQQQQQQQQQEQQRRFFQEQQBQQRRK  751

Qy   576  NEPAP--QQQQQPTKSYPRARKSRQGSTGSSPSGPQ-----------GISG------  613
               :  |||||                               :
Db   752  EEQQQQIQRQQQHLQQLQQQMEQQHVATAAPR-PQMHHPAPARLPTRMGGAAGHTALKS  810

Qy   614  --SKSFRPFAAVDEDSNINNAPEQTMTT---------TTTRTTVTQTTRDGGGVT  659
            ::    |||                    :::    :
Db   811  ELSEKFQMLRA------NNNSSMMRMSGTDLEGLADVLKSEITTSLSALVDTI-----VT  859

Qy   660  RVIKVVPHNAKLASENAARIFQSIQEERK  688
            :     |            :  :
Db   860  RFV----HQRRLFSKQADSVTAAAEQLNK  884


RESULT 23
ID   YQKA_SCHPO     STANDARD;      PRT;    964 AA.
AC   Q9UTW9;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Protein C1494.10 in chromosome III.
GN   SPCC1494.10 OR SPCC70.01.
OS   Schizosaccharomyces pombe (Fission yeast).
OC   Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC   Schizosaccharomycetales; Schizosaccharomycetaceae;
OC   Schizosaccharomyces.
OX   NCBI_TaxID=4896;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=972;
RX   MEDLINE=21848401; PubMed=11859360;
RA   Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,
RA   Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,
RA   Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,
RA   Collins M., Connor R., Cronin A., Davis P., Feltwell T., Fraser A.,
RA   Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,
RA   Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,
RA   James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,
RA   Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,
RA   Oliver K., O'Neil S., Pearson D., Quail M.A., Rabbinowitsch E.,
RA   Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,
RA   Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,
RA   Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,
RA   Woodward J., Volckaert G., Aert R., Robben J., Grymonprez B.,
RA   Weltjens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,
RA   Gabel C., Fuchs M., Fritzc C., Holzer E., Moestl D., Hilbert H.,
RA   Borzym K., Langer I., Beck A., Lehrach H., Reinhardt R., Pohl T.M.,
RA   Eger P., Zimmermann W., Wedler H., Wambutt R., Purnelle B.,
RA   Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaure V., Mottier S.,
RA   Galibert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,
RA   Lucas M., Rochet M., Gaillardin C., Tallada V.A., Garzon A., Thode G.,
RA   Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,
RA   Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,
RA   Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,
RA   Shpakovski G.V., Ussery D., Barrell B.G., Nurse P.;
RT   "The genome sequence of Schizosaccharomyces pombe.";
RL   Nature 415:871-880(2002).
RN   [2]
RP   SEQUENCE OF 420-596 FROM N.A. AND SUBCELLULAR LOCATION.
RC   STRAIN=968 h90;
RX   MEDLINE=20223868; PubMed=10759889;
RA   Ding D.-Q., Tomita Y., Yamamoto A., Chikashige Y., Haraguchi T.,
RA   Hiraoka Y.;
RT   "Large-scale screening of intracellular protein localization in living
RT   fission yeast cells by the use of a GFP-fusion genomic DNA library.";
RL   Genes Cells 5:169-190(2000).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic and nuclear.
CC   -!- SIMILARITY: Contains 1 LisH domain.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AL023794; CAA19351.1; -.
DR   EMBL; AL023776; CAA19308.1; -.
DR   EMBL; AB027952; BAA87256.1; -.
DR   GeneDB_SPombe; SPCC1494.10; -.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; ISS.
DR   GO; GO:0006350; P:transcription; ISS.
DR   InterPro; IPR006594; LisH.
DR   SMART; SM00667; LisH; 1.
DR   PROSITE; PS50896; LISH; 1.
KW   Nuclear protein.
FT   DOMAIN        34     66       LISH.
SQ   SEQUENCE   964 AA;  104591 MW;  BBCC13AD417D66F4 CRC64;

Query Match             4.2%;
Best Local Similarity   22.6%;
Matches   92;  Conservative   59;  Mismatches  169;  Indels   87;  Gaps   20;

Qy   234  QESRNRLYRDGGKTRLKDTDNGAEBSHLATRNHSQBGHGSP---EDIDNDREYSKSRACA  289
          :|||  ::  :             ::|:|  |      ::|
Db    60  RESKIQIPKSSSSTAFSPSNNNAPSPFPPKNSSL--ASPSKISESISGDRLYNHMSSAP  116

Qy   290  SLQQINE---EASDDVSDDSMVDSISSIDVSPDDV----VGILGQ--KRFWRARKAIA  338
          :::                   :||
Db   117  SPNKKEETNVVHANEDISLDKR-QSFGSSSLPPSEVSINVPEGFIVEWFNIFWDVFSARV  175

Qy   339  NQQRVFAVQLFE--LHRLIKVQKLIAASPDLLLDEISIPLGKVSAKSYPVK-KLLPSEFL  394
            ::   :   :||||:
Db   176  SRVNSTPIQLYDPSTQROMARPMSNLQASQPVPSSTFSRSAVVPNPSLPLNPSVLQGQVM  235

Qy   395  VKPLPHVVKQRGDSEKTD---QHKME---SSAENVVGRLSNQGHHQQSNYMPFANNPP  448
             :|:  :||:
Db   236  NNPTIPKGTPSTSIEGAKTSIPPSHAMQNPHNSFPASADRLQKNHPVQSSNFNPYTPAPS  295

Qy   449  ASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPM  508
             :||
Db   296  ITVPPN---------------------YIPNTAM-----MGPSYSSFGDTD-  320

Qy   509  VMPQYHP-GMGFPPPG--NGYFPPYGMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQ  565
             ||:   |||:        |
Db   321  --PRTYPAGMGPNTAARNGFYPP---TPAQIH-QLKAQQQHLQRQSKQMSE-PAPINMK  373

Qy   566  NTQQQQQ----RSDNEPAPQQQQQPTKSYPRARKSRQGSTGSSPSG  607
             ||:           ||  :||||:
Db   374  SNKDQQLQYVDFRGVGSGADLQKQQWNKS--------TSAERLQPNG  412


RESULT 24
ID   SMF1_HUMAN     STANDARD;      PRT;   1902 AA.
AC   O14497; Q9UPZ1;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
```

```
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   SWI/SNF-related, matrix-associated, actin-dependent regulator of
DE   chromatin subfamily F member 1 (SWI-SNF complex protein p270) (B120).
GN   SMARCF1 OR C1ORF4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., AND MUTAGENESIS.
RX   MEDLINE=20221560; PubMed=10757798;
RA   Dallas P.B., Pacchione S., Wilsker D., Bowrin V., Kobayashi R.,
RA   Moran E.;
RT   "The human SWI-SNF complex protein p270 is an ARID family member with
RT   non-sequence-specific DNA binding activity.";
RL   Mol. Cell. Biol. 20:3137-3146(2000).
RN   [2]
RP   SEQUENCE OF 1-1175 FROM N.A.
RX   MEDLINE=98094256; PubMed=9434167;
RA   Takeuchi T., Chen B.-K., Qiu Y., Sonobe H., Ohtsuki Y.;
RT   "Molecular cloning and expression of a novel human cDNA containing CAG
RT   repeats.";
RL   Gene 204:71-77(1997).
RN   [3]
RP   SEQUENCE OF 1-1132 FROM N.A.
RA   Takeuchi T., Misaki A.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: BINDS DNA NON-SPECIFICALLY.
CC   -!- SUBUNIT: Part of the SWI-SNF complex.
CC   -!- SUBCELLULAR LOCATION: Nuclear (Potential).
CC   -!- TISSUE SPECIFICITY: Highly expressed in spleen, thymus, prostate,
CC       testis, ovary, small intestine, colon, and pbl, and at a much
CC       lower level in heart, brain, placenta, lung, liver, skeletal
CC       muscle, kidney, and pancreas.
CC   -!- SIMILARITY: Contains 1 ARID domain.
CC   -!- CAUTION: Ref.2 sequence differs from that shown due to a
CC       frameshift in position 1132.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF265208; AAF75765.1; ALT_INIT.
DR   EMBL; AB001895; BAA23269.1; ALT_FRAME.
DR   EMBL; AB024075; BAA83073.1; ALT_SEQ.
DR   EMBL; AB024059; BAA83073.1; JOINED.
DR   EMBL; AB024060; BAA83073.1; JOINED.
DR   EMBL; AB024061; BAA83073.1; JOINED.
DR   EMBL; AB024062; BAA83073.1; JOINED.
DR   EMBL; AB024063; BAA83073.1; JOINED.
DR   EMBL; AB024064; BAA83073.1; JOINED.
DR   EMBL; AB024065; BAA83073.1; JOINED.
DR   EMBL; AB024066; BAA83073.1; JOINED.
DR   EMBL; AB024067; BAA83073.1; JOINED.
DR   EMBL; AB024068; BAA83073.1; JOINED.
DR   EMBL; AB024069; BAA83073.1; JOINED.
DR   EMBL; AB024070; BAA83073.1; JOINED.
DR   EMBL; AB024071; BAA83073.1; JOINED.
DR   EMBL; AB024072; BAA83073.1; JOINED.
DR   EMBL; AB024073; BAA83073.1; JOINED.
DR   EMBL; AB024074; BAA83073.1; JOINED.
DR   Genew; HGNC:11110; SMARCF1.
DR   MIM; 603024; -.
DR   GO; GO:0005634; C:nucleus; NAS.
DR   GO; GO:0003677; F:DNA binding; NAS.
DR   InterPro; IPR001606; ARID.
DR   InterPro; IPR008938; ARM.
DR   Pfam; PF01388; ARID; 1.
DR   SMART; SM00501; BRIGHT; 1.
```

```
KW   DNA-binding; Nuclear protein.
FT   DOMAIN        96     99       POLY-GLN.
FT   DOMAIN       178    184       POLY-GLN.
FT   DOMAIN       615    618       POLY-SER.
FT   DOMAIN       631    740       ARID.
FT   DOMAIN       944   1021       GLN-RICH.
FT   MUTAGEN      690    690       W->A: PARTIAL LOSS OF DNA-BINDING
FT                                 ACTIVITY. WHEN A.A. 713 ALSO MUTATED,
FT                                 COMPLETE LOSS OF ACTIVITY.
FT   MUTAGEN      713    713       Y->A: PARTIAL LOSS OF DNA-BINDING
FT                                 ACTIVITY. WHEN A.A. 690 ALSO MUTATED,
FT                                 COMPLETE LOSS OF ACTIVITY.
FT   CONFLICT      27     27       D -> G (IN REF. 1).
FT   CONFLICT      51     51       V -> M (IN REF. 1).
FT   CONFLICT     349    349       Q -> S (IN REF. 2).
FT   CONFLICT     368    368       G -> GG (IN REF. 3).
FT   CONFLICT     374    374       P -> S (IN REF. 2).
FT   CONFLICT     393    393       P -> L (IN REF. 2).
FT   CONFLICT     489    502       MANMPPQVGSGMCP -> NGQYATSGWVRDVS (IN REF.
FT                                 2).
FT   CONFLICT     634    634       E -> G (IN REF. 2).
FT   CONFLICT     924    924       P -> S (IN REF. 2).
FT   CONFLICT    1016   1016       Q -> P (IN REF. 2).
FT   CONFLICT    1033   1033       Q -> L (IN REF. 2).
SQ   SEQUENCE    1902 AA;  205946 MW;  9B70A46F81062EAA CRC64;
```

```
Query Match             4.2%;  Score 154.5;  DB 1;  Length 1902;
Best Local Similarity  21.7%;  Pred. No. 0.91;
Matches  124;  Conservative    62;  Mismatches  203;  Indels  183;  Gaps  28;

Qy        173 GSKQNVCLATCSKPE--------------VRDQVKANARSGGFVISLLDVSVTEEIDLEKSASSHDR 224
                : : ||                          : |:
Db        569 GTMANNSAGMAASPEMMGLGDVKLTPATKMNNKADG-------------TPKTESKSKKSSSS- 618

Qy        225 VNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLA-TENHSQEGHGSP----EDIDND 279
                           :  |   |||         :|:||
Db        619 -------TTTNEKITKLYELGGEPERK---MWVDRYLAFTEEKAMGMTNLPAVGRKPLDIY 669

Qy        280 REYSKSRACASLQQINE------EASDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRAR 334
                                        : ::  :|::  ||
Db        670 RLYVSVKEIGGLTQVNKNKKWRELATNLNVGTSSSAASSL------------------ 709

Qy        335 KAIANQQRVFAVQLFELHRLIKVQKLIAASPDILLDEISFLGKVSAKSYPVKKLLPSEFL 394
                :  |                                            ||
Db        670 ---------KAIANQQRVFAVQLFELHRLIKVQKLIAASPDILLDEISFLGKVSAKSYPVKKLLPSEFL 394

Qy        710 --------KKQYIQCLYAFE----CKIERGEDPPPDI-----FAAADSKKSQP---------K 746
                           |: ||                        ||
Db        710 ...

Qy        395 VKPPLPHVVVKQRGDSEKTDQHKMESSAE------------------------NVVG- 427
                ::  :                                                   |
Db        747 IQPPSPAGSGSMQG--PQTPQSTSSSMAEGGDLKPPTPASTPHSQIPPLPGMSRSNSVGI 804

Qy        428 -RLSNQGH--HQQSNYMPFANNPPASPAPN----------GYCFPQPPPSGNHQQ--WLIPV 475
                 || ::                                     :: |:  :
Db        805 QDAFNDGSDSTFQKRNSM--TPNPGYQPSMNTSDMMGRMSYEPNKDPYGSMRKAPGSDPF 862

Qy        476 MSPSEGLIYKPH------------------PGMAHTG----------HYGGYGHYMPTPMVMPQYHPG 516
                ||   ||                    :                :
Db        863 MSSGQG----PNGMGDPYSRAAGPGLGNVAMGRQHYPYGGPYDRVRTEPGIGPEGNMS 918

Qy        517 MGFPPPGNGYFPPYGMPTMNP----------------------YCSSQQQQQQPNEQM-NQF---------G 559
                ||                                        :||||||||
Db        919 TGAFQP-------NLMPS--NPDSGMYSPSRYPPQQQQQQQRHDSYGNQFSTQGTPSG 968

Qy        560 HPGNLQNT--QQQQQRSDNEPAPQQQQQPTKSYPPRARKSRQGSTGSS----------PSGPQ 609
                ::                  ||||| :
Db        969 SPFPSQQTTMYQQQQQNYKRPMDGTYGPPAKRHEGEMYSVPYSTQGQPQQQQLPPAQPQ 1028

Qy        610 GISGSKSFRPFAAVDEDSNINNAPEQTMTTTT 641
                :
Db       1029 PASQQQAAQPSPQQDVYNQYGNAYPATATAAT 1060
```

RESULT 25
SAS_DROME

```
ID   SAS_DROME      STANDARD;     PRT;   1693 AA.
AC   Q04164; Q960M6; Q9VI73;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Putative epidermal cell surface receptor precursor (Stranded at second
DE   protein).
GN   SAS OR CG2507
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM LONG).
RX   MEDLINE=9290115; PubMed=1339334;
RA   Schonbaum C.P., Organ E.L., Qu S., Cavener D.R.;
RT   "The Drosophila melanogaster stranded at second (sas) gene encodes a
RT   putative epidermal cell surface receptor required for larval
RT   development.";
RL   Dev. Biol. 151:431-445(1992).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196606; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C.J., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D.A., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J.H., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   REVISIONS, AND ALTERNATIVE SPLICING.
RX   MEDLINE=22426069; PubMed=12537572;
RA   Misra S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA   Hradecky P., Huang Y., Kaminker J.S., Millburn G.H., Prochnik S.E.,
RA   Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA   Bettencourt B.R., Celniker S.E., de Grey A.D.N.J., Drysdale R.A.,
RA   Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.Q.,
RA   Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA   Lewis S.E.;
RT   "Annotation of the Drosophila melanogaster euchromatic genome: a
RT   systematic review.";
RL   Genome Biol. 3:RESEARCH0083.1-RESEARCH0083.22(2002).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM SHORT).
RC   STRAIN=Berkeley; TISSUE=Embryo;
RX   MEDLINE=22426066; PubMed=12537569;
RA   Stapleton M., Carlson J.W., Brokstein P., Yu C., Champe M.,
RA   George R.A., Guarin H., Kronmiller B., Pacleb J.M., Park S., Wan K.H.,
RA   Rubin G.M., Celniker S.E.;
RT   "A Drosophila full-length cDNA resource.";
RL   Genome Biol. 3:RESEARCH0080.1-RESEARCH0080.8(2002).
CC   -!- FUNCTION: Vital for larval development.
CC   -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=Long;
CC         IsoId=Q04164-1; Sequence=Displayed;
CC       Name=Short;
CC         IsoId=Q04164-2; Sequence=VSP_004071;
CC   -!- TISSUE SPECIFICITY: Expressed in most, if not all, ectodermal
CC       tissues which produce a cuticle.
CC   -!- DEVELOPMENTAL STAGE: Throughout development.
CC   -!- SIMILARITY: Contains 1 fibronectin type III domain.
CC   -!- SIMILARITY: Contains 2 VWFC domains.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; M68866; AAA28879.1; -.
DR   EMBL; AE003672; AAN13346.1; -.
DR   EMBL; AE003672; AAF54052.2; -.
DR   EMBL; AY051979; AAK93403.1; -.
DR   FlyBase; FBgn0002306; sas.
DR   GO; GO:0016324; C:apical plasma membrane; NAS.
DR   GO; GO:0005887; C:integral to plasma membrane; NAS.
DR   GO; GO:0004872; F:receptor activity; NAS.
DR   GO; GO:0002168; P:larval development (sensu Insecta); IMP.
DR   InterPro; IPR008957; FN_III-like.
DR   InterPro; IPR003961; FN_III.
DR   InterPro; IPR001007; VWF_C.
DR   Pfam; PF00041; fn3; 1.
DR   SMART; SM00060; FN3; 3.
DR   SMART; SM00214; VWC; 4.
DR   PROSITE; PS01208; VWFC_1; 1.
DR   PROSITE; PS50184; VWFC_2; 1.
KW   Receptor; Developmental protein; Signal; Transmembrane; Glycoprotein;
KW   Repeat; Alternative splicing.
FT   SIGNAL        1     41
FT   CHAIN        42   1693       Putative epidermal cell surface receptor.
FT   DOMAIN      106    106       FIBRONECTIN TYPE-III.
FT   DOMAIN      109    109       EXTRACELLULAR (POTENTIAL).
FT   TRANSMEM   1636   1656       POTENTIAL.
FT   DOMAIN     1657   1693       CYTOPLASMIC (POTENTIAL).
FT   DOMAIN       51    377       THR-RICH.
FT   DOMAIN      534    901       CYS-RICH.
FT   DOMAIN      997   1273       GLU/PRO-RICH.
FT   DOMAIN      663    708       VWFC 1.
FT   DOMAIN      828    902       VWFC 2.
FT   DOMAIN     1411   1501       POTENTIAL.
FT   CARBOHYD    106    106       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    109    109       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    330    330       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    473    473       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    538    538       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    622    622       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    685    685       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    827    827       N-LINKED (GLCNAC...) (POTENTIAL).
FT   CARBOHYD    846    846       N-LINKED (GLCNAC...) (POTENTIAL).
```

```
FT   CARBOHYD       929    929       N-LINKED (GLCNAC. . . .) (POTENTIAL).
FT   CARBOHYD       939    939       N-LINKED (GLCNAC. . . .) (POTENTIAL).
FT   CARBOHYD      1323   1323       N-LINKED (GLCNAC. . . .) (POTENTIAL).
FT   CARBOHYD      1419   1419       N-LINKED (GLCNAC. . . .) (POTENTIAL).
FT   CARBOHYD      1517   1517       N-LINKED (GLCNAC. . . .) (POTENTIAL).
FT   VARSPLIC       930   1274       Missing (in isoform Short).
FT                                   /FTId=VSP_004071.
FT   CONFLICT       591    591       V -> L (IN REF. 1).
FT   CONFLICT      1509   1509       V -> G (IN REF. 1).
SQ   1693 AA;  185253 MW;  DA50F9677F41DC4 CRC64;

Query Match             4.2%;  Score 153;  DB 1;  Length 1693;
Best Local Similarity   19.7%;  Pred. No. 0.95;
Matches   95;  Conservative   52;  Mismatches  169;  Indels  166;  Gaps   20;

Qy    157 APSSHHSIRFQEVNQTGSKQNVCLATCSK-PEV-RDQVKANARSGGFVISLDVSVTEEI- 213
Db    919 ATSTHSSIPANETTTTATANKSTSIPSKVPQIKKDEEKRPPASGAFYPTLDGKPPKSIG  978

Qy    214 ---DLEKSASSHDRV------NDYNASLRQESRNRLYRDGGKTRLKD---TDNGA 256
Db    979 GLGIFEKPEKPEKAHKKVQHQQQQHQQEQQQHQNDVIFDGDRTEEQEEPLPPNGGF 1038

Qy    257 ESHLATENHSQEGHGSP---------EDIDNDREYSKSRACASLQQINEEASDDVSDD 305
Db   1039 VPFQFGQQHPQPHLGPYGFYNPVKPVYEDYNPYFPY--------DINPNGTPQ---- 1084

Qy    306 SMVDSISSIDVSPDDVVGIIGQKRFWRARKAIANQQRVFAVQLFELHRLLIKVQKLIAASP 365
Db   1085 ---GKPPPVPTSQSDLFNILGAEQ----                                 1105

Qy    366 DLLLDBISFLGKVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKTDQHKMESSAENV 425
Db   1106 ---------PGHPVH--PGH--GGPPRIHPGQTQKDNHNLGPQVRIEQILQ--       1143

Qy    426 VGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYK 485
Db   1144 ------HLQQT------VPGG---PPPPPHQQHQS-LTPQLHPQQQQISQ          1178

Qy    486 PHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGY--------FPPYGMMPTI 536
Db   1179 QHP---------GHYVPI------VHSGVPPPPPGHGIAIVDGQTVAYESYPVIPGL   1220

Qy    537 MNPYCSSQQQQQQPNEQMNQFGHPG------NLQNTQQQQQRSDNEPAPQQQQQPTKS 589
Db   1221 GVPQ-HHPQQHQTTPQQHLQQTILPSSTTSGLSTQASEHSLHQNQGKLAKQQQSGANNL 1279

Qy    590 YP 591
Db   1280 QP 1281
```

Search completed: July 29, 2004, 08:22:44
Job time : 23 secs

OM protein - protein search, using sw model

Copyright (c) 1993 - 2004 Compugen Ltd.

GenCore version 5.1.6

Run on: July 29, 2004, 08:20:25 ; Search time 79 Seconds
(without alignments)
2775.761 Million cell updates/sec

Title: US-10-719-885-2
Perfect score: 3669
Sequence: 1 MKRGKDEKILEPMFPRLHV........AARIFQSIQEERKRYDSSKP 695

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 100 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	773.5	21.1	760	10	Q9SNQ6	Q9snq6 oryza sativ
2	704.5	19.2	765	10	Q942Q5	Q942q5 oryza sativ
3	409	11.1	623	10	Q9LU26	Q9lu26 arabidopsis
4	200.5	5.5	1208	5	Q86HF5	Q86hf5 dictyosteli
5	192	5.2	1503	5	Q9V6P4	Q9v6p4 drosophila
6	190	5.2	1099	4	Q8IW32	Q8iw32 homo sapien
7	189.5	5.2	802	3	Q07034	Q07034 saccharomyc
8	189.5	5.2	1500	5	Q8IHH0	Q8ihh0 dictyosteli
9	187.5	5.1	1165	5	Q86HN1	Q86hn1 drosophila
10	187	5.1	498	5	Q7YXC7	Q7yxc7 caenorhabdi
11	186	5.1	682	5	Q9U010	Q9u010 drosophila
12	186	5.1	745	5	061458	061458 drosophila
13	186	5.1	745	5	Q8IMG5	Q8img5 drosophila
14	186	5.1	771	5	062609	062609 drosophila
15	184.5	5.0	756	11	Q9QY60	Q9qy60 mus musculu
16	183.5	5.0	960	5	Q8I9J8	Q8i9j8 drosophila
17	182.5	5.0	3938	11	O88778	O88778 rattus norv
18	182	5.0	671	5	Q8IMG4	Q8img4 drosophila
19	182	5.0	697	5	Q8IMG6	Q8img6 drosophila
20	180	4.9	1594	5	Q9V6W7	Q9v6w7 drosophila
21	179.5	4.9	1749	4	Q86IN4	Q86in4 dictyosteli
22	178.5	4.9	1162	4	Q96JK9	Q96jk9 homo sapien
23	178.5	4.9	697	5	O77069	077069 drosophila
24	178.5	4.9	1024	10	Q93YP4	Q93yp4 arabidopsis
25	178.5	4.9	1239	5	O01505	001505 caenorhabdi
26	178.5	4.9	3942	11	O88737	088737 mus musculu
27	178	4.9	1015	5	Q8MRN6	Q8mrn6 drosophila
28	178	4.9	1420	5	Q9Y0Y8	Q9y0y8 drosophila
29	177.5	4.8	2090	5	Q9VHK3	Q9vhk3 drosophila
30	177	4.8	748	10	Q9T0G5	Q9t0g5 arabidopsis
31	177	4.8	771	5	O76259	076259 drosophila
32	176.5	4.8	2210	4	Q7Z3E7	Q7z3e7 homo sapien
33	176	4.8	1023	5	Q9LX42	Q9lx42 arabidopsis
34	175.5	4.8	887	3	Q8J0T7	Q8j0t7 candida alb
35	175	4.8	382	11	Q8R1B3	Q8r1b3 mus musculu
36	174.5	4.8	496	5	Q8VZR8	Q8vzr8 arabidopsis
37	174	4.7	564	5	Q9VJ62	Q9vj62 drosophila
38	174	4.7	2201	5	Q86NN1	Q86nn1 drosophila
39	174	4.7	2529	5	Q9VWC0	Q9vwc0 drosophila
40	173.5	4.7	1724	5	P91019	P91019 caenorhabdi
41	173	4.7	2957	5	O61845	061845 caenorhabdi
42	172.5	4.7	759	12	Q9LLF0	Q9llf0 white spot
43	172	4.7	1359	5	Q9VX26	Q9vx26 caenorhabdi
44	171	4.7	789	5	Q9U2W6	Q9u2w6 caenorhabdi
45	171	4.7	1758	5	O16625	016625 caenorhabdi
46	170.5	4.6	1183	11	Q7TSH6	Q7tsh6 mus musculu
47	170	4.6	945	5	Q7YU70	Q7yu70 drosophila
48	170	4.6	1608	5	Q8MQW4	Q8mqw4 drosophila
49	170	4.6	2531	5	Q9VW47	Q9vw47 drosophila
50	169.5	4.6	588	3	Q8TGE7	Q8tge7 aspergillus
51	169	4.6	584	5	Q9NJB0	Q9njb0 spodoptera
52	168.5	4.6	527	5	Q9VVD7	Q9vvd7 drosophila
53	168.5	4.6	1193	4	Q8N3I8	Q8n3i8 homo sapien
54	168	4.6	749	5	Q9BIW6	Q9biw6 drosophila
55	168	4.6	1256	5	Q22126	Q22126 caenorhabdi
56	168	4.6	2584	5	Q9TW88	Q9tw88 caenorhabdi
57	168	4.6	2606	5	Q21920	Q21920 caenorhabdi
58	168	4.6	2678	5	Q9NDS4	Q9nds4 caenorhabdi
59	166.5	4.5	1399	4	O43793	043793 homo sapien
60	166.5	4.5	1441	4	O43792	043792 homo sapien
61	166.5	4.5	1838	5	Q960G3	Q960g3 drosophila
62	166	4.5	2063	5	Q86B73	Q86b73 drosophila
63	165.5	4.5	386	13	Q7ZXW1	Q7zxw1 xenopus lae
64	165.5	4.5	577	5	Q86IP6	Q86ip6 dictyosteli
65	165.5	4.5	740	3	Q96U76	Q96u76 neurospora
66	165.5	4.5	832	5	Q7YYP6	Q7yyp6 cryptospori
67	165	4.5	729	5	Q17796	Q17796 caenorhabdi
68	165	4.5	780	11	Q9EQV7	Q9eqv7 rattus norv
69	165	4.5	918	10	O82206	082206 arabidopsis
70	165	4.5	1833	5	Q9VUH6	Q9vuh6 drosophila
71	164.5	4.5	979	10	Q8RXR5	Q8rxr5 arabidopsis
72	164.5	4.5	673	5	Q8MNL6	Q8mnl6 dictyosteli
73	164.5	4.5	1396	5	Q8T0T0	Q8t0t0 drosophila
74	164	4.5	1318	16	Q7UG24	Q7ug24 rhodopirell
75	163.5	4.5	2310	5	Q9W2U7	Q9w2u7 drosophila
76	163.5	4.5	5322	5	Q9VPL9	Q9vpl9 drosophila
77	163	4.4	688	5	Q9I7J5	Q9i7j5 drosophila
78	163	4.4	690	5	Q9NH72	Q9nh72 drosophila
79	163	4.4	690	5	Q9VFT4	Q9vft4 drosophila
80	163	4.4	1006	11	Q62901	Q62901 rattus norv
81	163	4.4	1128	5	Q9VZ60	Q9vz60 drosophila
82	163	4.4	1499	4	Q86UU0	Q86uu0 homo sapien
83	163	4.4	2465	5	Q8IQ71	Q8iq71 drosophila
84	163	4.4	2473	5	Q86IP9	Q86ip9 dictyosteli
85	162.5	4.4	1293	5	Q8INQ0	Q8inq0 drosophila
86	162.5	4.4	1475	5	Q9N4G4	Q9n4g4 caenorhabdi
87	162.5	4.4	1725	11	Q80U46	Q80u46 mus musculu
88	162	4.4	749	5	Q9W384	Q9w384 drosophila
89	162	4.4	858	11	Q99N37	Q99n37 rattus norv

```
 90    161.5    4.4    530    10    Q9LYK5          Q9lyk5 arabidopsis
 91    161.5    4.4   1169    11    Q8CHB3          Q8chb3 mus musculu
 92    161.5    4.4   1263    11    Q80Y82          Q80y82 mus musculu
 93    161.5    4.4   3498     5    Q20497          Q20497 caenorhabdi
 94    161      4.4   1083    16    O86637          O86637 streptomyce
 95    161      4.4   1146     3    Q8X0L3          Q8x0l3 neurospora
 96    161      4.4   1322     5    Q9VVH2          Q9vvh2 drosophila
 97    160.5    4.4    508    10    Q9M0H8          Q9m0h8 arabidopsis
 98    160.5    4.4   1441     4    O00150          O00150 homo sapien
 99    160      4.4    699    10    Q8LQU0          Q8lqu0 oryza sativ
100    160      4.4    752    11    Q8R506          Q8r506 rattus norv
```

ALIGNMENTS

RESULT 1

```
Q9SNQ6
ID   Q9SNQ6       PRELIMINARY;       PRT;   760 AA.
AC   Q9SNQ6;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to Arabidopsis thaliana chromosome II BAC T19L18.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare;
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 6, PAC
RT   clone:P0535G04.";
RL   Submitted (AUG-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP000399; BAA83571.1; -.
DR   Gramene; Q9SNQ6; -.
DR   GO; GO:0006364; P:rRNA processing; IEA.
DR   InterPro; IPR000238; Rib_bind_facta.
DR   PROSITE; PS01319; RBFA; 1.
SQ   SEQUENCE   760 AA;  82378 MW;  253BF51F93D35670 CRC64;
```

```
Query Match             21.1%;  Score 773.5;  DB 10;  Length 760;
Best Local Similarity   32.1%;  Pred. No. 3.4e-46;
Matches  261; Conservative 105; Mismatches 257; Indels 191; Gaps 36;

Qy      4 GKDEE-KILEPMFPRLHVNDADK-GGPRAPPRNKMALYEQLSIPSQRF----GDHGTMNS 57
            ||:|| ||||||||||||||:|  ||:|||||||||||||||:|:|      ||||||:
Db     12 GKEAKGKVMGPLFPRLHVNDAAKGGGPRAPPRNKMALYEQFTVPSHRFSGGGGGVGGS 71

Qy     58 RSNNTSTLVHPGPSSQPCGVERNL------------SVQHLDSSAANQATEKFVSQMSF 104
            :||||:||||||||| |||||:|             ::|||::||| : |:|:||:|:
Db     72 PAHSTSAASQSQSQSQVYGRDSSLFQPFNVPSNRPGHSTEKINSDKINKKISSSRKEIGM 131

Qy    105 MEN---------VRSSAQHDQR-----------KMVREEEDFAVPVYINSRRSQSHGRTK 144
            |:|         :|::||| |           |:|::::::|:|:|:|:||: :::|
Db    132 LSSQTKGMDIYASRSTAEAPQRRAENTIKSSSGKRLADDDEFMVPSVFNSRFPQYSTQEN 191

Qy    145 SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV-CLATCSKPEVRDQVK-ANARSGGFV 202
            |||||||||||||||||||||||||||||||||  ||:|||:||||:|| ||||:||:|
Db    192 AGVQ-DQSTPLVAANPHKS-----PSTVSKSSTKCYNTVSKKLERIHVSDVKSRTPLKD 244

Qy    203 ISLDVSVI-EEIDLEKSASSHDRVNDYNASLRQESRN-RLYRDGKTR--LKDTD---NGA 256
            ||:|:|:  :|:||::|:|:|| |:|:| ::|:|:|  ||:|| ::   |||:    ||
Db    245 KEMEAAQTSKNVEVEKSSSFHASKDMF----ESRHAKVYPKMDKTGIINDSDEPHGGN 298

Qy    257 ESHLATENHS---------QBGHGSPEDIDNDREY-------SKSRACASLQQIN 295
            |:||:|||           :|:||:|||:|:||         |||:|||:||:|
Db    299 SGHQATSRNGGSMKFQNPPMRRNEISSNPSSENTDRHYNLPQGGIEETGTKRKRLLEQHD 358

Qy    296 EEASDDVS---------DDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQ 341
            || ||||         ||||| |: ::| ||||||||||||||||||||||||
```

```
Db    359 AEKSDDVSRLLEQHDAENIDDVSDSSVECITGWEISPDKIVGAIGTKHFWKARRAIMNQQ 418
            ||                       |||||| |||:|:||:|:|:|:||||||||:|:||

Qy    342 RVFAVQLFELHRLIVKVQKLIAASPDLLDEISFLGKVSAKSYPVVKKLLPSEFLVKPPLPH 401
            |||||||||||||||||||||||||||||||:||||:|:||||||:||||||:|||||:||

Db    419 RVFAVQVFELHKLIVKVQKLIAASPHVLIESDPCLGNALLGS---KNKLVEENLKAQPLLV 475
            ||||||:||||:|||||||||||||::| :||  ||:|||||    ||||||:|||:|||

Qy    402 VVVKQRGDSEKTDQHKMESSAENV------VGRLSNQGHHQQSNYMPFAN--NPPA 449
            |||:||| |::|:|||:|:|:| |       ||:|::|:|::|||:|||   |||

Db    476 ATI-----DDVEPSLQQPEVSKENTEDSPPSHDTGLGSGQRDQAATNGVSKSNRRATPVA 531
            ||     |||::||:::|:|::::| |||::|:|:||:|:|:|||:|:||:||||||:|

Qy    450 SP-APNGYCFPPQPPSGNHQQWLIPVMSPSEGLIYKPHPG----------------M 490
            || ||:||||||:|||:|:|  |:|:|||||||:||||:|

Db    532 SDNKQNNWGVQLPP----QNQWLVPVMSPLEGLVIKPYSGPCPPAGSILAPFYANCTPL 587
            | ::|:||:|:|||    :|:|:|||||||||:|:||:||  ||||:||:||||||||

Qy    491 AHTGHYGGYYGHYMPTPMVMPQYHPQMGFPP--PGNGYFPPYGMPTIMNPYCSSQQQQQ 548
            |:||:||: |:|:|||||||||||||||||   |||||||||||| ||||:||:|||||

Db    588 SLPSTAGDFMNSAYGVPMPHQPHPQHMGAPGPPSMPMN-YFPPSI--PVMNP 635
            |:||:|||:|:||: :|||: || |:|| |  |:|:| |||||   |:|||

Qy    549 QQPNEQMNQFGHPGNLQ--NTQQQQQRSDNEPAPQQQQQPTKSYP---RARKSR 597
            |:|:|:||| |:| :|   ::||||::|:||| ||||||| ||:    :|||||

Db    636 TAPAPVVEQGRHPSMPQPYGNFEQQSWISCN---------MSHPSGIWRFHASRDSEA 684
            |||| ||::| ||: || ||:|:||  :|          ||:| ||| |::||:|:|

Qy    598 QGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTTRTTVTQTTRDGGG 657
            ||::|:||:||:|:||||||||:| |:|||:||||||||:|||||||||| |||:|||||

Db    685 QASSASSPFDRFQCSGS--GPVSAPPTVSAQNNQPQP------------SYSSRD-N 726
            ||:|||||||:| ||||   ||:|||||:|:|:||          |||||: |

Qy    658 VTRVIKVPHNAKLASENAARIFQSIQEERKRYD 691
            |||||||||| |||:|||||||||||:|||:||

Db    727 QTNVIKVPHNSRTASHSAARIFRSIQMERQRDD 760
            |:|||||||: |:|:||||||:|||| ||:|||
```

RESULT 2

```
Q94Q05
ID   Q94Q05       PRELIMINARY;       PRT;   765 AA.
AC   Q94Q05;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   P0697C12.15 protein.
GN   P0697C12.15.
OS   Oryza sativa (Rice).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC   Ehrhartoideae; Oryzeae; Oryza.
OX   NCBI_TaxID=4530;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=cv. Nipponbare.
RA   Sasaki T., Matsumoto T., Yamamoto K.;
RT   "Oryza sativa nipponbare(GA3) genomic DNA, chromosome 1, PAC
RT   clone:P0697C12.";
RL   Submitted (FEB-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AP003296; BAB64680.1; -.
DR   Gramene; Q94Q05; -.
SQ   SEQUENCE   765 AA;  82653 MW;  5CE1C45576D4D83C CRC64;
```

```
Query Match             19.2%;  Score 704.5;  DB 10;  Length 765;
Best Local Similarity   29.5%;  Pred. No. 2.6e-41;
Matches  244; Conservative 113; Mismatches 263; Indels 207; Gaps 38;

Qy      4 GKDEE--KILEPMFPRLHVNDADK-GGPRAPPRNKMALYEQLSIPSQRF-GDHGTMNSR 58
            ||:||  ||||||||||||||:|  ||:|||||||||||||||:||:| ||||:||:|

Db      8 GKEVEERGKVMGPLFPRLHVNDAAKGGGPRAPPRNKMALYEQFTVPSHRFSGGGGGALASA 67

Qy     59 SNNTSTLVHPGPSSQPCGVERNL------------SVQHLDSSAAN 92
            :||||:|||||||| |||||:|             ::|||::||

Db     68 RGSLARSTSAASQSQVYGCDMPLFEPFNVPSNGPGQSVEKMNSNVRQINGSRKDSGML 127

Qy     93 ---QATEKFVS---QMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRTK 144
               :|||:||   |||:|:|||:||||| |||::::::|:|:|:|:||::::|:|:

Db    128 STQPKGIDKYGSGSRAECAPQQRVEKGIKSSSSGRKLADDDEFIVPSVFSARFPQYSTKER 187

Qy    145 SGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSK-----PEVRDQVKANAR 197
            |||||||||||||||||||||||||||||||||||||||||     ||||:||||||
```

```
Query Match             11.1%;  Score 409;  DB 10;  Length 623;
Best Local Similarity   22.8%;  Pred. No. 1.7e-20;
Matches  177;  Conservative  98;  Mismatches  98;  Indels 195;  Gaps  33;

Qy     5 KDEEK--ILEPMPPRLHVNDADKGGPRAPPRNKMALYEQLS----IPSQRFGDHGTMN   56
         ||||| :||:|:||||:||.: .::          ||::||:   ||::|.|: |:|:
Db    59 KDEAKRITIPPLFPRVHVNDTGRGG----------LSQQFDGKTMSLVSSKRPNLPSPTN 108

Qy    57 SRSNNTSTL--VHPGPSSQPC--GVERNLSVQHLDSSAANQATEKFVS----QMSFMEN  107
         :|||:||:|  ||||::|:    |.||:|:|.:|: :||:|:. |:.|     :|:
Db   109 NISDSLSTFSLSLPPPNNARLIDGPEKNQFSPIYNTKFEGKLNKKGLNYTSPKGSSVTN  168

Qy   108 VRSSA--QHDQRKMVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIR  165
         |::.   :|:|:|.|::::|:|:|:||||:|:||.:|| |::|||:|||:|||:|||:||
Db   169 TKPSSIKQNEYLKNLTLSLDSIKSPIVHSE----------IDPQANTDL-------SLQ  210

Qy   166 FQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSASSHDRV  225
         |.||:||.||:||.|:||.||:||||||:|:||||||||||||||||||||||||||||
Db   211 F--------------GGEAVVGSKILLSE-----------------------------RL  235

Qy   226 NDYNASLRQESRNRLIYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKS  285
         :::::.:||:||||:|||||||||||||||||||||||||||||||||||||||||||||
Db   236 EDEN--------------QNGSPNVMKTQSYRR----NFAEFNNETQ-KKP          267

Qy   286 RACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARK-----------  335
         :|.||:|||::|:||||||||||||||||||||||||||||||||||||
Db   268 KTLPRREQV--------ASNCSAIESLSGISASSYDIARVIGEKRFWKMRTYMIKFVSSSH  320

Qy   336 ---------------ALANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLG      376
                        |||:||||||||||||||:|||:|||||||||||||||||||||
Db   321 FSYDPNFITLPYFLLFFIVDIFSQQKIFAGQVFELHRLIMVQKMVAKSPNLFLES----  375

Qy   377 KVSAKSYPVKKLLPSEFLVKPPLPHVVVKQRGDSEKT--DQHKMESSAENVVGRLSNQGH  434
         |::||:|||||||||:|||||||.||:|||.||:|    .::||:::||:||.|.|:.|
Db   376 ---------------KLNGVKHGTMRSSHQLAMAASKV--RKPNTEN              405

Qy   435 HQQSNYMPFANNPASPAPNGY---------------CFPPQ-PPPSGNHQQWLI        473
         |::|||||||||||:||.||                |||||  ||||||::|||
Db   406 HK-------------PVPEEYPEHMKPKLPLPSISKELVTPIWPQQLLPPGN--QWLV  449

Qy   474 PVMSPSEGLIYKPHPGMAHTGHYGYGGYYGHYMPTP-------MVMPQY---HPGMGPP-  520
         ||:||||||||||||||:||.|||||.:||:|:|:        |.||||    ||:||
Db   450 PVITDSDGLVYKPFPG--------------PCPPSSAAFMVPVYGQDSLETPFRFPVSS  494

Qy   521 PPGNGYFPPYGMMPTIMNPYCSSQQQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAP  580
         |||||||||||:||:|||||||||||||||.||:||||:. :::      .|:.:.||
Db   495 PFSHSYFPPPNARTTV-----------------DQTNPFG------QFQRWSNTSSH    528

Qy   581 QQQQQP--TKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTM  637
         ||||||  |||||||||||:||||||||||.||||||||||||||||||||||||||
Db   529 MTQAIPFSLKKKSQESNDSDIHGSTASSP-------------------------PEKHK  561

Qy   638 TTTTTTTRTTVTQTT---RDGGGVTRVIKVVPHNAKLASENAARIFQSIQEERKRYD  691
         |||||||||:|:||     |||||||||||||||||||||||||||||||||||||
Db   562 LEVLFLFFTEPTHQTDEYKQKQQPMLRAIKAVPHNSTSASESAARIFRFIQEERRDSD  619
```

```
Db   188 AGVQES-TPLVALSPHKSP--PAVSKSPTK---CNTVSXNLERINVSDVKSRGSQKDK  241
Db   198 SGGFVISL-DVSVTEIDLEKS------NDYN-ASLRQESRNRLIYRDG            244
Qy   242 ETGPAQTLKNVEVEHFSSFFASKDMFGSKHAKVCPKTGTINDLDEPHLENSEHQATSRNG  301
Db   245 GKTRLKD--------TDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQ    293
Qy   302 SSVKFQNPVVRRNTISAKPSPGIEN--TNGHCNLPQGGLKEAGTKRKRLEA------Q  351
Qy   294 INEEASDDVSDDSMVDSISSIDVSPDDVVGILGQKRFWRARKAIANQQRVFAVQLFELHR  353
Qy   352 DNAEKIDDLSDSS-VBCITAWEISPDIVGAIGAKHFWKARRAIINQQRVFAAQVPELHK  410
Qy   354 LIK-------VQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLP--KKKMA         390
Qy   411 LVKVSSWIEFGVEYENNQNLYLTYKQVQKLIAASPHVLIEGDPCLGNALLAS--KKKMA  467
Qy   391 SEFLVKPPLPHVVVKQRGDSEKTDQH--KMESSAENVVGRLSN--QGHHQQSNYMPFA   444
Qy   468 EENLKAQP---VLVATNDDVQPSLQEPELSKENSEENPFSPRDTAPVSGHHDQTAKIGAS  524
Qy   445 NN-------PPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYKPHPG-------  489
Qy   525 KSNLRATPVASDNRQNNCGVQLQPP---QNQWLIPVMSPSEGLVYKPYSGCPPPAGSILA  581
Qy   490 ------------MAHTGHY-GGYYGHYMPTPMPQYHPGMGFPPPGNGYFPPYGMMPTIM  537
Qy   582 PFYANCTPLRLPSTTGDFMNSAYG--VPIPH-QPQHGAPGTPTMPMNYFPPFSV--PVM  636
Qy   538 NPYCSSQQQQQQQPNEQMNQFGHPGNLQ---NTQQQQQRSDNEPAPQQQQQQPTKSYP--  591
Db   637 NPVALASAVEQGR-----------HPSMPQPYGNLEQHSRMSCN-                676
Qy   592 -RARKSR----QGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTRT  646
Db   677 WRFHASRDSEAQASSASSPFDRLQCGGSGPVSAFP-----------TASAQNTQP      720
Qy   647 TVTQTTRDGGGVTRVIKVVPH-NAKLASENAARIFQSIQEERKRYDS            692
Db   721 QPSSGGSRD-NQTNVIRVIPHNNSQTASESAARIFRSIQMERQQDDS            765

RESULT 3
Q9LU26
ID   Q9LU26;
AC   Q9LU26;
DT   01-OCT-2000 (TrEMBLrel. 15, Created)
DT   01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT   01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DbDbj|BAA83571.1.
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RA   Sato S., Nakamura Y., Kaneko T., Kato T., Asamizu E., Tabata S.;
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Columbia;
RX   MEDLINE=20277480; PubMed=10819329;
RA   Nakamura Y.;
RT   "Structural analysis of Arabidopsis thaliana chromosome 3. I. Sequence
RT   features of the regions of 4,504,864 bp covered by sixty P1 and TAC
RT   clones.";
RL   DNA Res. 7:131-135(2000).
DR   EMBL; AB023045; BAB01726.1; -.
SQ   SEQUENCE   623 AA; 69949 MW; DF6E3FF7B4E5D5D5 CRC64;
```

RESULT 4
Q86HF5
```
ID   Q86HF5;                      PRELIMINARY;      PRT;   1208 AA.
AC   Q86HF5;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to Homo sapiens (Human). HIWI.
OS   Dictyostelium discoideum (slime mold).
OC   Eukaryota; Mycetozoa; Dictyosteliida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910;
```

```
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., Abril J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC115682; AAO52645.1; -.
DR   InterPro; IPR003100; PAZ.
DR   InterPro; IPR003165; Piwi.
DR   Pfam; PF02170; PAZ; 1.
DR   Pfam; PF02171; Piwi; 1.
DR   PROSITE; PS50821; PAZ; 1.
DR   PROSITE; PS50822; PIWI; 1.
SQ   SEQUENCE   1208 AA;  137164 MW;  B20B9B69B71CE89C CRC64;

Query Match             5.5%;  Score 200.5;  DB 5;  Length 1208;
Best Local Similarity  24.5%;  Pred. No. 2.5e-05;
Matches   81;  Conservative   28;  Mismatches  100;  Indels  121;  Gaps  13;

Qy        386 KKLLPSEFLVKPPLPHVVVKQRGDSEKTDCHKMESSAENVVGRLSNQGHHQQSNYMPFAN 445
                :|||:|
Db          8 RDLLPSD-----------DDGGDLENR-KSRREDNIRHPQQQQQNYQQQNYQQQQN 54

Qy        446 -----NPPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGG 498
                        :|||:|  | |||:|:  | ||:|||:|||:|||:| :|
Db         55 YQQQPYDQQSQQSSQGGYYQPPPPPSGYYQ------SP-------PPYSRAGY--D 96

Qy        499 YYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQF 558
                     :||:| |||: |||:| | ||:|||:|||:|||:|||||||||||||||||
Db         97 PYGYYAPP-----PQHQPNSYYPQQPS--FDPYFQQQYQPPPPPQQQQQYQPQSQQQQY 150

Qy        559 ---------------GHPGNLQNTQQQQRSDNEPAPQQ-QQPTKSYPRARKSR 597
                               |||::||:||:|:| |:||||:|| |||:|:|:|||||
Db        151 QPQQSHGYDYQQSYQGYQNQQQPQQSQENQSQQQQSQPQRPQQQSYQSYQQQQPQQ 210

Qy        598 Q----------------------GSTGSSPSGPQ------------GISGS------ 614
                                       ||:|:|:|||              |:|:|
Db        211 QQQDIFQVPQDRNRSYRDSGGDRGNGSSGNDNYGYQRGRGRDNNNNNNGGGGSGSGDRYG 270

Qy        615 ----KSFRPFAAVDEDSNINN 631
                    ||:|::|:|:||::||
Db        271 SGSGGSGSGDRYSGRGRRDFGDRDNNNNNNNN 300
```

```
RESULT 5
Q9V6P4
ID   Q9V6P4;        PRELIMINARY;        PRT; 1503 AA.
AC   Q9V6P4;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
DE   CG4744 protein.
GN   CG4744.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Wan K.H., Doyle C., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
```

```
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhao Q., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
DR   EMBL; AE003818; AAF58379.1; -.
DR   FlyBase; FBgn0033834; CG4744.
SQ   SEQUENCE   1503 AA;  171359 MW;  4302EA538BFCBE2C CRC64;

Query Match             5.2%;  Score 192;  DB 5;  Length 1503;
Best Local Similarity  22.9%;  Pred. No. 0.00014;
Matches  151;  Conservative   74;  Mismatches  244;  Indels  190;  Gaps  35;

Qy          4 GKDEEKILEPMFP------------RLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSR 58
                                       :||:|  |:|  :|:||
Db        398 GMDGNRMME--LPQSYDYRVHQYD-----RFFHLNQ-----RQMS----NVGDY---ER 437

Qy         59 SNNTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRK 118
                :| | ||:|:|:|:| :| ::|  ||:|: | ::| | | | | :|::|||:|:|||:
Db        438 SNR---------------MPLDLDRFKREPHSESSFLEAPNELSLGQLSVEEPVLKS----EDEN 483

Qy        119 MVREEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV 178
                    :||:||:|  |:|:|: || :| ||||||:|  | ||: :|||:|:|:||||:|
Db        484 ISNREMAMKIAV-LDIKDDKALHSTGDGSKEKDT------------------LEEVKLEGLHEEL 530

Qy        179 CL--ATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASSHDRVNDYNASLRQES 236
                  :|:|  :  | ::|:|:  || ||:|:||:||:||:| | |: |:|:||:|:||:|:|
Db        531 KLDEESLVKPESKLDSEVSCINGKF------EDRSEMEVDNKPLGPVK-REAKAELINQM 583

Qy        237 RNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSK-SRACASLQQIN 295
                || |  |:|| |:|||:|||| :||:| |::|:|||:||:|:|:|||:| ||:|||:||:|
Db        584 GNKLKNDGSTLVAK----SSAESPLKEDVPPNLARRSELNVSTGTDLSAVSKKQNSLKKRK 640

Qy        296 EEASDDVSD--DSMVD----------SISSIDVSPDDVVGILGQKRFWRARKAIANQQRV 343
                :|: || || |:|:||          :|:|:| |:|: | |:||:||||:|:|:||:|:|
Db        641 SSATSDDSDKDHSNPDCDTECKSEDAEKSETDAEKDD------------VKKLLPSEFLVK 677

Qy        344 FAVQLFELHRLLIKVQKLIAASPDLLLDEIS-FLGKVSAKSYP------VKKLLPSEFLVK 396
                ||::|  ||||:||:|| |::| ::||||:| |:|||:|||| |       |||||||||||
Db        678 ---SIVKVNIKLKLQBPVEEKSE--CEEISKVVGKGAAAASAPKVNPDAVVLLANELSMD 732

Qy        397 PPLPHVVKQRG----------DSEKTDQHKMESSAENVVGRLS--NQGHH 435
                ||||||| |||          |||:|| ||||:|:|:|||:||  :|||:
Db        733 ------KDRGKRQARHRYRRQNRRSKRSLEDNDQHESDDKQHAVSKRLAPLLDDFED 784

Qy        436 QQSNYMPFANNPPASPAPNGYCFPQPPPSGNHQQWLLIPVMSPSEGLIYKPHPGMAH--- 492
                | | |||| |:||| | ||:| |||:|||:|||: ||:|||:|||:|||:|||:|:| |
```

```
Db   785  HNGNRVGFGN------LGNG-MLIPDAEYSGERS-----------YPSHEMLAHFDE  823
                ::||     :||: | :: | :                    :||||
Qy   493  --TCHYG-GYYGHYMPTPMVMPQYH--PGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ  547
                 ::| :::|||:.:|:| |:::|                     :|||
Db   824  LETAHDGDSMVPIYDPTKVVVPMVNQRPNGGF------QGGQPVQMQ--FSQQQP  870

Qy   548  QQQPNEQMNQFGH-----PGNLQNTQQQQRSDNEPAPQQQQ-QPTKSYPRARKSRQGS  600
                :|||                 : |||                  |:::|
Db   871  QQQMQQQMQQ-QHMQQQQPPQGPGQQQQQQQQQQQQVQQPQQHDLSRQRHQSS  928

RESULT 6
Q8IW32
ID    Q8IW32           PRELIMINARY;        PRT;    1099 AA.
AC    Q8IW32;
DT    01-MAR-2003 (TrEMBLrel. 23, Created)
DT    01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    Similar to R3H domain (binds single-stranded nucleic acids)
DE    containing.
OS    Homo sapiens (Human).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX    NCBI_TaxID=9606;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    TISSUE=Uterus;
RA    Strausberg R.;
RL    Submitted (DEC-2002) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; BC041093; AAH41093.1; -.
DR    GO; GO:0003676; F:nucleic acid binding; IEA.
DR    InterPro; IPR001374; R3H.
DR    Pfam; PF01424; R3H; 1.
DR    SMART; SM00393; R3H; 1.
SQ    SEQUENCE  1099 AA;  120696 MW;  67782B9163702174 CRC64;

Query Match             5.2%;   Score 190;  DB 4;  Length 1099;
Best Local Similarity  18.9%;   Pred. No. 0.00012;
Matches 162;  Conservative 102;  Mismatches 293;  Indels 298;  Gaps 37;

Qy    42  LSIPSQRFGDHGTMNSRSNNTSTIVH------------PGPSSQPCGVERNLSVQ---------  84
               |:||   |.|   :|:||:                      | ::::|     ||::|
Db    56  LQRPLQSFGQTG-KRSKSSSKLKIVRSLAVCEESPPPAPEISQENQEKIQIQLTQSFEK  114

Qy    85  ----HLDSSAANQATEKFVSQM----------------SFMENVRSSAQHDQRKMVRE  122
                  |||:|:|:|::||:|                  ::|::||:|||:|  |::|:|
Db   115  EEKPSKDEAEKEKASDKLPRKMLSRDSSQEYTDSTGIDLHEFLVNTLKNNPRDRMMLLKL  174

Qy   123  EEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEV----------NQTGS  174
               :||.:| |||||:||||||:|||||:.| ||  |:|::|:::                :::
Db   175  EQE--ILDFIGNNES----------PRKKFPPM---TSYHRMLLHRVAAYFGLDHNVDQS  219

Qy   175  KQNVCLATCSKPEVRDQV-----------KANARSGGFVISLDVSVTEEIDLE-----------K  217
                |::  | ::|||:|:                   :::    ||:|:|| :|::||:
Db   220  GKSVIVNKTSNTRIPDQKFNEHIKDDKGEDFQKRYILLKRDNSSFDKDDNQMRIRLKDRR  279

Qy   218  SASSHDRVNDYNASLRQESRNRLYRDGG----------KTRLKDTD-----------  253
               ::|::|||:|::|::||:|||||:||||                |||:||:|:
Db   280  SKSIEEREEY-----QRARDRIFSQDSLCSQENYIIDKRLQDEDASSTQQRRQIFRVNK  334

Qy   254  ---NGAESH-LATEN----HSQEGHGSPEDIDNDREYSK-----------  284
                   |||:| ||||:     | :| :|:   :|::|||||
Db   335  DASGRSTNSHQSSTENELKYSEPRPWSSTDSDSLRNLKPAVTKASSFSGISVLTRGDSS  394

Qy   285  --SRACASLQQINEEASDDVSDDSMVDSISSID-----------VSPDDV  321
                  :|||:|||::||:||::|||:|:|:|:|                 ||:||:
Db   395  GSSKSIGRLSKTGSESSGSVG--SSTGSLSHIQQPLPGTALSQSSHGAPVVPTVSTHSS  452

Qy   322  V----GIIGQKRFWRARKAIANCQRVFAVQLFELHRLIKVQKLI---AASPDLLLDEISF  374
                     :::|  |||  | | :|| |:|:|||||:||:|:|          :|:::|:|||:
Db   453  LSFDGGLNGQ-----------VASPSTSFFLLPLEAAGIPPGSIIINPQTGQPFINPDGSPV  503

Qy   375  LGKVSAKSYPVKKLLPSEFLVKPPLP-----------HVVVKQRGDSEK  412
               |  :::||:|::|| | :| :::                 |:
Db   223  DRVNDYNASLRQESRNRLYRDGGKTRLKDTDNG-----------AESHLAT  262
```

```
Db   504  VYNPEMTQQPVRSQVPGP--FQPPLPAPPQQPAANHIFSQDNLGSQFSHMSLARQPSADG  561
                 :::|:|              :||||                  :: ::
Qy   493  --TCHYG-GYYGHYMPTPMVMPQYH--PGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ  547

Db   464  QGQPVQMQ--FSQQQP  870

Qy   610

Db   562  SDPHAAMFQSTVVL---QSPQQSGYIMTAAPPHPPP----PPPPPLPLPGQ  610

Qy   465  -----SGNHQQWLIPVMSP--SEGLIYKPHGM--AHTGHYGGYYGHYMPTM  508

Db   611  PVPTAGYPASGH------PVSQPVLQQGYIQQPSPQMPACYCAPGHYHSSQPQYRPVPS  664

Qy   509  VMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ-----QQPNEQMNQFGHPGN  563
                ::|:         :|||:|:|            ||:|::|       |||::||::|||
Db   665  V--HYNSHLNQPLPQAQQTGYQVIP-----NQQQNYQGIVGVQQPQSLVSGQPNS  715

Qy   564  LQN---TQQQQRSDNEPAPQQQQQPTKSYPRARKSRQGTGSSPSGPQG  610
                :::      | :|                                   :
Db   716  IGNQIQGVVIPYTSVPTYQVSLPQGSGIPHQTYQQPVMFP----NQSNQGSMPT-  766

Qy   611  ISGSKSFRPFAAVDEDSNINNA------------PEQTMTTTTTT  653
                :::::::|:|                        ::||:||:|::
Db   767  -TGMPVYYSVIPPGQQNNLSSSVGYLQHPGSEQVQPRTTSPCSSQQLQGHQCTAGPPPP  825

Qy   654  DGGGVTRVIKVVPHN  668
               :|::|::|: |:|||
Db   826  PGGGMVMMQLSVPNN  840

RESULT 7
Q07034
ID    Q07034           PRELIMINARY;        PRT;    802 AA.
AC    Q07034;
DT    01-NOV-1996 (TrEMBLrel. 01, Created)
DT    01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT    01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE    RNA binding protein.
OS    Saccharomyces cerevisiae (Baker's yeast).
OC    Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC    Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX    NCBI_TaxID=4932;
RN    [1]
RP    SEQUENCE FROM N.A.
RA    Sugimoto K., Matsumoto K., Kornberg R.D., Reed S.I., Wittenberg C.;
RT    "HMD1 of the yeast Saccharomyces cerevisiae encodes a putative
RT    RNA/single strand DNA-binding protein, whose overexpression reduces
RT    the expression of a G1 cyclin, CLN3.";
RL    Submitted (AUG-1994) to the EMBL/GenBank/DDBJ databases.
DR    EMBL; D37935; BAA07154.1; -.
DR    GO; GO:0003676; F:nucleic acid binding; IEA.
DR    InterPro; IPR000504; RNA_rec_mot.
DR    Pfam; PF00076; rrm; 1.
DR    SMART; SM00360; RRM; 1.
DR    PROSITE; PS50102; RRM; 1.
SQ    SEQUENCE  802 AA;  90437 MW;  2F320C882E0D063F CRC64;

Query Match                  5.2%;   Score 189.5;  DB 3;  Length 802;
Best Local Similarity       20.2%;   Pred. No. 8.5e-05;
Matches 128;  Conservative 73;  Mismatches 216;  Indels 217;  Gaps 29;

Qy    81  LSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKM-----------VREEEDFAVPV  130
               |||:|:||||||:|||||||||||:|||||||||::::           |::|:||||:
Db   253  LSVPHDSTSTTASASATSGARSNDQRKPPLSDAQRMRFPRADLSKPITEEHDRYAA  312

Qy   131  YINSRRSQSHGRTK-----------SGIEKEKHTPMVAPSSHHSIRFQEVN  170
               |:|:|:|                    ::|||:|::|| ::||:::|::|:
Db   313  YL------HGENKITEMHNIPPKSRLFIGNLPLKIVSKEDLFRIFSPYGH--IM  358

Qy   171  QTGSKQNVCLATCSKPE-VRDQVKANARSGGFVISLDVSVTEEIDLEKSASS------H  222
                :: ::| : : :: |  :| ||:||||||||:|||||||:||:|:||||      |
Db   359  QINIKNAFGFIQFDNPQSVRDAIECE----------SQEMNFGKKLILEVSSNARPQFDHG  410

Qy   223  DRVNDYNASLRQESRNRLYRDGGKTRLKDTDNG-----------AESHLAT  262
               | :: :  :  ::|      ::        |:
```

```
Db   411 DHGTNSSTFISSAKRPQTESGD--MYNDDNGAGYKKSRRHTVSCNIFVKRTADRTYAI 468
Qy   179 CL--ATCSKPEVRDQVKANARSGGFVISLDVSVTEIDLEKSASSHDRVNDYNASLRQES 236
                :   :: :                     :::  :::

Db   263 E--NHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDD 320
Qy   531 KLDEESLVKPESKLDSEVSCINGKF-----EDRSEMEVENKPLGPVK-REAKAELINQM 583
              :    :: :            : :  :             :

Db   469 EVFNRFRDGTG---LETMIFLKPRMELG-KLINDAAYNGVWGVLVLNKTHNVDV---- 519
Qy   237 RNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDINDREYSK-SRACASLQQIN 295
          : :  ::     : :                     :  ::: :: :

Db   321 VVGILGQKFWRARKAIANQQRVFAVQLFELRHRLIKVQKLIAASPDLLLDEISFLGKVSA 380
Qy   584 GNKLKNDGSTLVAK--SSAESPLKEDVPPNLARRSELNVSTGTDLSAVSKKQNSLKKRK 640
               :           :    :            :::    :   :

Db   520 ------QTFYKSQGET--------------------KFDEYIS1SAD---DAVAIFNNIK- 551
Qy   296 EEASDDVSD--DSMVD--------------------SISSIDVSPDDVVGILGQKRFWRARKAIANQQRV 343
                                            :: :   :

Db   552 -----------NNRRNSRPTDYRAMS------------HQQNIY 572
Qy   641 SSATSDDSDKHDSNPCDTECKSEDAEKSETDAEKDD-------- 677
                                           :::  :

Db   441 ---MPFANNPPASPAP----NGYCFPP----QPPPSGNHQQWLIPVMSPSEGLIYKPH 487
Qy   344 FAVQLFELHRLIKVQKLIAASPDLLLDEIS-FLGKVSAKSYP-----VKKLLPSEFLVK 396
             :                          :  :::   :

Db   573 GAPPLPVPNGPAVGPPPQTNYYQGYSMPPPQQQQQQPYGNYG---MPPPSHDQG--YGSQ 627
Qy   397 ------DSEKTDQHKMESSAENVVGRLS---NQGHH 435
                 :    :         :    :  :

Db   488 PGMAHTGHYGGYHGYHMPTPMVMPQYHPGMGF------PPPGNGYFPP------- 529
Qy   733 ----------KDRGKRQARHRYRRQNRRSKRSLEDNDQHESDDKQHAVSKRLAPLLDDFED 784
                                                     :

Db   628 PPIPMNQSYGRYQTSIPPPP---PQQQIPQGYGRYQAGFPPQ----PPSQTPMDQQQLLS 680
Qy   436 QQSNYMPFANNPPASPAPNGYCFPQPQPPSGNHQQWLIPVMSPSEGLIYKPHPGMAH--- 492
           : :                 :          :::: :

Db   530 --YGMMPTIMNPYCSSQQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPT 587
Qy   785 HNGNRVGFGN------LGNG-MLTPDAEYSGGRS----------YPSHEMLAHFDE 823
                  :                 :                 ::

Db   681 AIQNLPPNVVSNLLSMAQQQQQQPHAQQQLV---GLIQSMQGQ----APQQQQQQL 729
Qy   493 --TGHYG-GYYGHYMPTPMVMPQYH--PGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ 547
            :::                     :::  :: :

Db   588 KSYPRARKSRQGSTGSSPSGPQGIGSKSFRPFA 621
Qy   824 LETAHDGDSMVPIYDPTKVVVPMVNQRPNGGF------QGQPVQMQ--FSQQQP 870
              ::                          :

Db   730 GGYSSMNSSPPPMSTNYNG-QNISAKPSAPPMS 762
Qy   548 QQQPNEQMNQ--FGHPGNLQNTQQQQRSDNEPAPQQQQQVQQP-QQHDLSRQRHQSS 600
            ::                                 :

Qy   871 QQQMQQQMQQQHMQQQPQPQQPQQPQQQQQQQQQQQQQQQVQQP-QQHDLSRQRHQSS 925
```

```
RESULT 8
Q8IHH0
ID   Q8IHH0        PRELIMINARY;        PRT;  1500 AA.
AC   Q8IHH0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   AT02321p.
GN   CG4744.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Dresnek D., Farfan D., Frise E.,
RA   George R., Gonzalez M., Guarin H., Kronmiller B., Li P., Liao G.,
RA   Miranda A., Mungall C.J., Nunoo J., Pacleb J., Paragas V., Park S.,
RA   Patel S., Phouanenavong S., Wan K., Yu C., Lewis S.E., Rubin G.M.,
RA   Celniker S.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BT001259; AAN71015.1; -.
DR   FlyBase; FBgn0033834; CG4744.
SQ   SEQUENCE  1500 AA;  170871 MW;  1AD8D6686FEC90C0 CRC64;

Query Match          5.2%;   Score  189.5;   DB  5;  Length 1500;
Best Local Similarity  22.6%;  Pred. No. 0.00021;
Matches   148;  Conservative  74;  Mismatches  247;  Indels  187;  Gaps  33;

Qy     4 GKDEEKILEPMFP------RLHVNDADKGGPRAPPRNKMALYEQLSIPSQRFGDHGTMNSR 58
           : ::                                   :
Db   398 GMDGNRMME--LPQSYDYIRVHQYD-----RFHPLNQ------RQMS----NVGDY----ER 437

Qy    59 SNNTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRK 118
                 :              :      :   :
Db   438 SNR----MPLDLDRFKREPHSESSFLEAPNELSLGQLSVEEPVLKSEDEN--- 483

Qy   119 MVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNV 178
           : :             :: :
Db   484 -ISDHEMAMKIAVLDIKDDKALHSTGDGESKEKDT-------LEEVKLEGLHEEL 530
```

```
RESULT 9
Q86HN1
ID   Q86HN1        PRELIMINARY;        PRT;  1165 AA.
AC   Q86HN1;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to PH (Pleckstrin homology) domain.
OS   Dictyostelium discoideum (Slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., Abril J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116957; AAO52542.1; -.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   InterPro; IPR001849; PH.
DR   InterPro; IPR000219; RhoGEF.
DR   InterPro; IPR000306; Znf_FYVE.
DR   Pfam; PF01363; FYVE; 1.
DR   Pfam; PF00169; PH; 1.
DR   Pfam; PF00621; RhoGEF; 1.
DR   SMART; SM00064; FYVE; 1.
DR   SMART; SM00233; PH; 1.
DR   SMART; SM00325; RhoGEF; 1.
DR   PROSITE; PS50010; DH_2; 1.
DR   PROSITE; PS50003; PH_DOMAIN; 1.
DR   PROSITE; PS50178; ZF_FYVE; 1.
```

SQ SEQUENCE 1165 AA; 129282 MW; 9F0F5558BE251A73 CRC64;

Query Match
Best Local Similarity 19.3%; Score 187.5; DB 5; Length 1165;
Matches 125; Conservative 101; Mismatches 243; Indels 177; Gaps 28;

```
Qy    115 DQRKMVREEDFA------------VPVVINSRRSQSHGRTKSGIEKEKHTPMVAPSHHSIRF 166
              |::::|||   ::|         :::  ||
Db    469 DKNRVQESSDITNSQLNADEVPVWIPDKRA-----TKCMFCNDGFT--IINRRHHCRNC 521

Qy    167 QEVNQTGS-----------KQNVCLATC----------SKPEVRDQVKANARSGGFVIS 204
                     ||      :  |  :||
Db    522 GKV-VCGSCSPHKRLIPHIKKNKPVRVCLFCYDYIGLNEKESQNHNQQSSSSSSSSSLH 580

Qy    205 LDVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGG------KTRLKDTDN- 254
          ||                 :  | ||         |         |
Db    581 INGAQRSHLDILNSI-----IGDRKRTL--SFSSLYGDGSIESNSTQPYNLTFNNNNNN 632

Qy    255 ----GAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMV 308
               ||
Db    633 GGSNGSSGSGNINININNINPNSSAGLSNKEHYS-----TFRKLKNKLSKDSSSSS- 684

Qy    309 DSISSIDVSPDDVVGILGQKRFWARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLL 368
                      :|
Db    685 --SSITITKE----------------KDRDFSRSQSDL--------------NLI 707

Qy    369 LDEISFLGKVSAKSYP-------VKKLLP-----SEFLVKPPLP--HVVVKQRGDSEKTD 414
                            ::    |                    ::
Db    708 GGGNSGSSGSGGSTPTGRPPLVVRGGIPLPTTQSHGNLRPMLPINSVILKPSNNNNNND 767

Qy    415 QHKMESS--AENVGRLSNQGHHQQSNYMRFANNPPAS-PAPNGYCFP-PQPPPSGNHQQ 470
                          |          |     ::
Db    768 QPKFQSPEFPQSKLSKTSPQSSSNNLNLITKSNKPPLSPPTPPHHNHPLPQTPPHHNHP- 826

Qy    471 WLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPY 530
                                                      ||||
Db    827 --LPPTPPQ------------------------HPLPPP---PPQHP---LPPIPQSHTPPQ 856

Qy    531 GMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQQQQPTKSY 590
          ||||||||||||||||||||||||||||||||
Db    857 HSLPSV------PQSHTPPSPSQQ--HPLIMQ-TKKVTRSTSQPPLPEQTHSSTTLQ 904

Qy    591 PRARKSRQGSTSGSPSGPGISGSKSFRPFAAVDEDSNINNAPEQTMTTTT-------- 643
                                                     :
Db    905 PSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSSISNPL 964

Qy    644 --TRTTVTQTTRDDGGGVTRVKVVPHNAKLASENAARIFQSIQEER 687
              ||
Db    965 GKSAPTITPSRRVPPPTQPRNSLENNTSASSPHISTEGQSLNNHQ 1010
```

RESULT 10
ID Q7YXC7 PRELIMINARY; PRT; 498 AA.
AC Q7YXC7;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE C. elegans PQN-58 protein (Corresponding sequence R10E12.1d).
GN PQN-58.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditina;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxID=6239;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology.";
RL Science 282:2012-2018(1998).
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN=Bristol N2;
RA Smith A.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DDBJ databases.
DR EMBL; Z29561; CAD91697.1; -
SQ SEQUENCE 498 AA; 55166 MW; 7B098D5B5F599A2D CRC64;

Query Match
Best Local Similarity 20.7%; Score 187; DB 5; Length 498;
Matches 126; Conservative 70; Mismatches 221; Indels 192; Gaps 28;

```
Qy     36 MALYEQLSIPSQRFGD-------------HGTMNSRSNNTSTLVHPGPSSQPCGVERNLSVQ 84
                                         |:||  |            ::::
Db      1 MATFGFLSAPLKSTNEVDLVKPLTSYIDNVYNTSDDNNRSDVAE---------------AVQ 46

Qy     85 HLD---SSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHG 141
                                                           ::::
Db     47 ELNKLRSKACCQPLDKHQSALDVL------TRYYDQLVAIENK---------IIISATQNPVVF 95

Qy    142 RTKSGIEKEKHTPMVAPSHHSIRFQE----------VNQTGSKQNVCLATCSKPEVRDQ 191
                                                    :|
Db     96 KWKDAFDKGSLFSSRASLSLSDGSFERAAVLFNIGSLMSQIGAAQQF--------HTDDE 147

Qy    192 VKANAR------SGGFVISLDVSVTEBIDLEKSAS-SHDRVNDYNASLRQESRNRLYRDGGK 246
                         |
Db    148 IKVSAXLFQQSAGVFARLRDVVLGMVQQEFPDLMPDTLAALSALMTAQAQEAIYIKGHK 207

Qy    247 TRLKDTDNGAESHLATENHSQEGHGSPEDI---------DND--------REYSKSR 286
                                                          ::
Db    208 DKMKATSMVKISAQVAEFYSEAQKMMSKDIVRGLWDKDWSAIVSGKNLAYQALAQYHQSE 267

Qy    287 ACASLQQINEEASDDVSDDSMVDSISSIDVS--------PDDVVGILGQKRFWARKAIANQQR 342
          ::
Db    268 VCGEARQIGEQLS----------RLAESLKLFDTAQKYLPRDITGI-------WDIYPSVSKAH- 314

Qy    343 VFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLPSEFLVKP-PLPH 401
                                                                :|
Db    315 -----------AAAKKD-------NDFIYHEKVS---DFRTLFTLPKAVLAKPTPM-- 349

Qy    402 VVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFP--- 459
          ::                                       |
Db    350 ----------------QTPMTPSFRGTNAAQSP------------ANAPRPPPPR-----PAA 380

Qy    460 ------PQPPPSGNHQQWLIPVMSPSEGLIYKPH----PGMAHTGHYGGYYGHYM---P 505
          ||
Db    381 PSVESPIPPPRTQQSMQATPGAPPQ----YNPYQQQQPQMQQFQQHPGYYQQPMPYGQP 436

Qy    506 TPMVMPQYHPGMGFP-PPGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNL 564
                                         |::
Db    437 QPMFQPYQPTFAAPYPTFPGAFSY------------QQQWP-QQQQQGFPPNP 479

Qy    565 QNTQQQQQR 573
          ||   |
Db    480 QFGQQNQQQ 488
```

RESULT 11
ID Q9U010 PRELIMINARY; PRT; 682 AA.
AC Q9U010;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE MEDEA-A (Fragment).
GN MED OR MEDEA OR CG1775.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98361878; PubMed=9694800;

RESULT 12

```
RESULT 12
O61458
ID   O61458      PRELIMINARY;      PRT;    745 AA.
AC   O61458;
DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MEDEA.
GN   MED OR CG1775.
RA   Xu X., Yin Z., Hudson J.B., Ferguson E.L., Frasch M.;
RT   "Smad proteins act in combination with synergistic and antagonistic
RT   regulators to target Dpp responses to the Drosophila mesoderm.";
RL   Genes Dev. 12:2354-2370(1998).
DR   EMBL; AF019754; AAC35437.1; -.
DR   HSSP; Q13485; 1YGS.
DR   TRANSFAC; T04379; -.
DR   FlyBase; FBgn0011655; Med.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IGI.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . . ; NAS.
DR   GO; GO:0008283; P:cell proliferation; IMP.
DR   GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR   GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
DR   GO; GO:0007476; P:wing morphogenesis; IMP.
DR   InterPro; IPR001132; Dwarfin.
DR   InterPro; IPR003619; Dwarfin_A.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF03165; MH1; 1.
DR   Pfam; PF03166; MH2; 1.
DR   SMART; SM00523; DWA; 1.
DR   SMART; SM00524; DWB; 1.
FT   NON_TER      682     682
SQ   SEQUENCE   682 AA;   72199 MW;  F4A38409DDA7F12A CRC64;

Query Match             5.1%;  Score 186;  DB 5;  Length 682;
Best Local Similarity  23.4%;  Pred. No. 0.00012;
Matches  111; Conservative  54; Mismatches 179; Indels 130; Gaps 24;

Qy  265 HSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI 324
Db   54 HRQGG------------ESEGFAKRAIESLVKKLKEKRDEL--DSLITAITTNGAHPSKCVTI 102

Qy  325 LGQKRFWRARKAIANQQ---RVFAVQLF--ELHR-LIKVQKLIAASPDLLLDEISF--- 374
Db  103 ---QRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHVKYCAFAFDLKCDSVCVNPY 159

Qy  375 ------------LGKVSAKSYPVKKLLPSEFLVKPLPHVVVKQRGDSEKTDQHKMESSAE 423
Db  160 HYERVVSPGIDLSGLSLQSGP-SRLVKDEYSAGPLVGSM------DIDGNDIGTIQHHPT 212

Qy  424 NVVGRLSNQGHHQ-QSNYMPFANNPPASPAPNGYCFP---PQPPPSGNHQQWLIP---- 474
Db  213 QMVGP--GGYGYPQGPSEYVGDA-NPMSAMFPTGRTIPKIEPQDGVAGSRGSWMVPPPPRL 270

Qy  475 --------VMSPSEGLIYKPHPGMAHTGHYGGYGHYMPTPMVMP---QYHPG--MGFPP 521
Db  271 GQPPQQQQQPQQ---QTPQPTQQQAQ-SQASAHSLPVPHGMPGPMNPGPVMAPPP 326

Qy  522 P------GNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQQQR 573
Db  327 PPQQAQNPQGNGVHHTQANSPTDPASALAMQQQQQQQQQQQQQQQQQQ---QQQQQQQ 376

Qy  574 SD------------NEPAPQQQQ------PTKSYPRARK-----SRQGSTG 602
Db  377 SGGVPNGSVNAGGAAAGGQYYSQPPPVSQMQGAGGGTSVAPSVHAQQNGYVSQPGSAG 436

Qy  603 SSPSPGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTTRRTVTQTTRDGG 656
Db  437 SAPVGGGVFGTAQPTP------------QQPQQPPTGVQANTGSAGAQAGAGGG 479
```

RESULT 13

```
RESULT 13
Q8IMG5
ID   Q8IMG5      PRELIMINARY;      PRT;    745 AA.
AC   Q8IMG5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Hudson J.B., Podos S.D., Keith K., Simpson S.L., Ferguson E.L.;
RL   Submitted (DEC-1997) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF039232; AAC09260.1; -.
HSSP; Q13485; 1DD1.
DR   FlyBase; FBgn0011655; Med.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IGI.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti. . .; IMP.
DR   GO; GO:0008283; P:cell proliferation; IMP.
DR   GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR   GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007179; P:TGFbeta receptor signaling pathway; IMP.
DR   GO; GO:0007476; P:wing morphogenesis; IMP.
DR   InterPro; IPR001132; Dwarfin.
DR   InterPro; IPR003619; Dwarfin_A.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF03165; MH1; 1.
DR   Pfam; PF03166; MH2; 1.
DR   SMART; SM00523; DWA; 1.
DR   SMART; SM00524; DWB; 1.
SQ   SEQUENCE   745 AA;   79071 MW;  316F0D71F32B7E9D CRC64;

Query Match             5.1%;  Score 186;  DB 5;  Length 745;
Best Local Similarity  22.8%;  Pred. No. 0.00014;
Matches  109; Conservative  56; Mismatches 174; Indels 140; Gaps 22;

Qy  265 HSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI 324
Db   28 HRQGG------------ESEGFAKRAIESLVKKLKEKRDEL--DSLITAITTNGAHPSKCVTI 76

Qy  325 LGQKRFWRARKAIANQQ---RVFAVQLF--ELHR-LIKVQKLIAASPDLLLDEISF--- 374
Db   77 ---QRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHVKYCAFAFDLKCDSVCVNPY 133

Qy  375 ------------LGKVSAKSYPVKKLLPSEFLVKPLPHVVVKQRGDSEKTDQHKMESSAE 423
Db  134 HYERVVSPGIDLSGLSLQSGP-SRLVKDEYSAGPLVGSM------DIDGNDIGTIQHHPT 186

Qy  424 NVVGRLSNQGHHQ-QSNYMPFANNPPASPAPNGYCFP---PQPPPSGNHQQWLIP---- 474
Db  187 QMVGP--GGYGYPQGPSEYVGDA-NPMSAMFPTGRTIPKIEPQDGVAGSRGSWMVPPPPRL 244

Qy  475 --------VMSPSEGLIYK---PH--PGMAHTGHYGGYGHYMPTPM 508
Db  245 GQPPQQQQQPQQTPQTQQQQAQSQAAAHSLPVPHGMPGMP-------GPMNPGPV 295

Qy  509 VMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQ 568
Db  296 MAPPPPQQAQNPQGNGVHHTQANSPTDPASALAMQQQQQQQQQQQQQ 345

Qy  569 QQQQRSD--------------NEPAPQQQQ------PTKSYPRARK-----SR 597
Db  346 QQQQQSGGVPNGSVNAGGAAAGGQYYGQPPPVSQMQGAGGGTSVAPSVHAQQNGYVSQ 405

Qy  598 QGSTGSSPSPGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTTRRTVTQTTRDGG 656
Db  406 PGSAGSAPVGGGVFGTAQPTP------------QQPQQPPTGVQANTGSAGAQAGAGGG 453
```

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG1775-PC.
GN   MED OR CG1775.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Celniker S.E., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
```

```
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003779; AAN14278.1; -.
DR   FlyBase; FBgn0011655; Med.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IGI.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti...; NAS.
DR   GO; GO:0008283; P:cell proliferation; IMP.
DR   GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR   GO; GO:0009880; P:leg morphogenesis (sensu Holometabola); IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
DR   GO; GO:0007476; P:wing morphogenesis; IMP.
DR   InterPro; IPR001132; Dwarfin.
DR   InterPro; IPR003619; Dwarfin_A.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF03165; MH1; 1.
DR   Pfam; PF03166; MH2; 1.
DR   SMART; SM00523; DWA; 1.
DR   SMART; SM00524; DWB; 1.
SQ   SEQUENCE 745 AA; 79085 MW; AA02AA6D8E0AB1A1 CRC64;

Query Match             5.1%;  Score 186;  DB 5;  Length 745;
Best Local Similarity   22.8%;  Pred. No. 0.00014;
Matches  109;  Conservative  56;  Mismatches  174;  Indels 140;  Gaps 22;

Qy      265  HSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI 324
                                                         :: ::
Db       28  HRQGG------------ESEGFAKRAIESLVKKLKEKRDEL--DSLITAITTNGAHPSKCVTI 76

Qy      325  LGQKRFWRARKAIANQQ--RVFAVQLF--ELHR-LIKVQKLIAASPDLLLDEISF---  374
                                         :  :: :     :: ::
Db       77  ---QRTLDGRLQVAGRKGPFHVIYARIWRWPDLHKNELKHVKYCAFAFDLKCDSVCVNPY 133

Qy      375  ------------LGKVSAKSYPVKKLLPSEFIVKPPLPHVVVKQRGDSEKTDQHKMESSAE 423
                                        :  :  ::::: :: ::
Db      134  HYERVVSPGIDLSGLSLQSGP-SRLVKDEYSAGPLVGSM-----DIDGNDIGTIQHHPT 186

Qy      424  NVVGRLSNQGHHQ-QSNYMPFANNPPASPAPNGYCFP---PQPPPSGNHQQWLIP------ 474
                                                       :::
Db      187  QMVGP-GGYGYPQGPSEYVGDA-NPMSAMFPTGRTIPKIEPQDGVAGSRGSWMVPPPPRL 244

Qy      475  --------------------VMSPSEGLIYK--PH--PGMAHTGHYGGYYGHYMPTPM  508
                                         ::          ::
Db      245  GQPPQQQQQPQQQQQTPQQTPQQQAQSQAAAHSLPVPHGMPGMP---------GPMNPGPV 295

Qy      509  VMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNBQMNQFGHPGNLQNTQ 568
                                                          :::::::
Db      296  MAPPPPPQQAQNPQGNGVHHTQANSPTDPASALAMQQQQQQQQQQQ---------Q 345

Qy      569  QQQQRSD------------------------NEPAPQQQQQ----------PTKSYPPARK---SR 597
                ::::
Db      346  QQQQQSGGVPNGSVNAGGAAAGQYYGQPPPVSQMQGAGGGTSVAPSVHAQQNGYVSQ 405

Qy      598  QGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQTMTTTTTTTTTRTTVTQTTRDGG 656
                                                        :::
Db      406  PGSAGSAPVGGGGVFGTAQPTP---------------QPPQQPPTGVQANTSAGAQAGAGGG 453

RESULT 14
062609
ID   062609     PRELIMINARY;       PRT;      771 AA.
AC   062609;
```

DT   01-AUG-1998 (TrEMBLrel. 07, Created)
DT   01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MED protein OR CG1775.
GN   MED OR MEDEA OR CG1775.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN bw;
RX   MEDLINE=98171467; PubMed=9502733;
RA   Das P., Maduzia L., Wang H., Finelli A., Cho S.-H., Smith M.,
RA   Padgett R.;
RT   "The Drosophila gene Medea demonstrates the requirement for different
RT   classes of Smads in dpp signaling.";
RL   Development 125:1519-1528(1998).
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=CN bw;
RX   MEDLINE=98171458; PubMed=9502724;
RA   Wisotzkey R.G., Mehra A., Sutherland D.J., Dobens L.L., Liu X.,
RA   Dohrmann C., Attisano L., Raftery L.A.;
RT   "Medea is a Drosophila Smad4 homolog that is differentially required
RT   to potentiate DPP responses";
RL   Development 125:1433-1445(1998).
RN   [4]
RP   SEQUENCE OF 27-771 FROM N.A.
RX   MEDLINE=98359837; PubMed=9693372;
RA   Inoue H., Imamura T., Ishidou Y., Takase M., Udagawa Y., Oka Y.,

RA   Tsuneizumi K., Tabata T., Miyazono K., Kawabata M.;
RT   "Interplay of signal mediators of decapentaplegic (Dpp): molecular
RT   characterization of mothers against dpp, Medea, and daughters against
RT   dpp.";
RL   Mol. Biol. Cell 9:2145-2156(1998).
RN   [5]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RA   Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
RA   Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,
RA   Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,
RA   Nunoo J., Pacleb J., Paragas V., Park S., Phouanenavong S., Wan K.,
RA   Yu C., Lewis S.E., Rubin G.M., Celniker S.;
RL   Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003779; AAF57172.1; -.
DR   EMBL; AF041439; AAC38972.1; -.
DR   EMBL; AF027729; AAC38971.1; -.
DR   EMBL; AF057162; AAC62005.1; -.
DR   EMBL; AY069496; AAL39641.1; -.
DR   HSSP; Q13485; 1DD1.
DR   TRANSFAC; T04379; -.
DR   FlyBase; FBgn0011655; Med.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IGI.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti.. .; NAS.
DR   GO; GO:0008283; P:cell proliferation; IMP.
DR   GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR   GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
DR   GO; GO:0007476; P:wing morphogenesis; IMP.
DR   InterPro; IPR001132; Dwarfin.
DR   InterPro; IPR003619; Dwarfin_A.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF03165; MH1; 1.
DR   Pfam; PF03166; MH2; 1.
DR   SMART; SM00523; DWA; 1.
DR   SMART; SM00524; DWB; 1.
SQ   SEQUENCE   771 AA;   81638 MW;   8E60A81E4F4946B3 CRC64;

Query Match   5.1%;   Score 186;   DB 5;   Length 771;
Best Local Similarity 22.8%;   Pred. No. 0.00014;
Matches 109;   Conservative 56;   Mismatches 174;   Indels 140;   Gaps   22;

Qy   265 HSQEGHGSPEDIDNDREYSKSRACASLQGINEEASDDVSDDSMVDSISSIDVSPDDVVGI 324
               ||||| |: :|::::| | | ||::| : |:::: ::|
Db   54 HRQQG------------ESEGAKRAIESLVKKLKEKRDEL--DSLITAITTNGAHPSKCVTI 102

Qy   325 LGQKRFWRARKAIANQQ---RVFAVQLF--ELHR-LIKVQKLIAASPDLLLDEISF--- 374
               || :| || :||:| :: | :: :|: |: |:| | : | ::|
Db   103 ---QRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHVKYCAFAPDLKCDSVCVNPY 159

Qy   375 ----------------LGKVSAKSYPVKKLLPSEFIVKPPLPHVVVKQRGDSEKTDQHKMESSAE 423
                               :| ||: |:| |: || : |::||:: |::   ::|: ::|
Db   160 HYERVVSPGIDLSGLSLQSGP-SRLVKDEYSAGPLVGSM------DIDGNDIGTIQHHPT 212

Qy   424 NVVGRLSNQGHHQ-QSNYMPFANNPPASPAPNGYCFP---PQPPPSGNHQQWLIP--- 474
               ||: |::|| :|:|: | : |: | | | :||:: |:
Db   213 QMVGP-GGYGYPQGPSEYVGDA-NPMSAMFPTGRTIPKIEPQDGVAGSRGSWMVPPPRL 270

Qy   475 ----------VMSPSEGLIYK----PH--PGMAHTGHYGGYGHYMPTPM 508
                         :|| |:|| | || ||:| |:|| | |:|:|:|
Db   271 GQPPQQQQQPQQQTPQTQQQAQSQAAAAHSLPVPHGMGMCMP-------- 321

Qy   509 VMPQYHPGMGFPPPGNGYFPPYGMPTIMNPYCSSQQQQQQPNEQMNQFGHPGNLQNTQ 568
               ||::|:| | ||| || | | | ||:|:| | : | :: |: : |:|:|::
Db   322 MAPPPPPQQAQNPQGNGVHHTQANSPTDPASALAMQQQQQQQQQQQQQQ----------Q 371

Qy   569 QQQQRSD------------------------------NEPAPQQQQQ 597
               ||: |: | |||
Db   372 QQQQQSGGVPNGSVNAGGGAAAGGQYYGQPPPVSQMQGAGGGTSVAPSVHAQQNGVVSQ 431

Qy   598 QGSTSGSSPSGPQGIGSGKSKFRPFAAVDEDSNINNAPEQTMTTTTTTTRTTVTQTTRDGG 656

```
Db       432  PGSAGSAPVGGGGVFGTAQPTP-----------QQPQQPPTGVQANTGSAGAQAGAGGG 479
              |||:|||:|:||:                              ::|                 ||

RESULT 15
Q9QY60
ID   Q9QY60;        PRELIMINARY;    PRT;    756 AA.
AC   Q9QY60;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Cg1/Xap80 protein.
GN   CG1/XAP80.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Platzer M., Zhao W., Herman G.E., Rosenthal A.;
RT   "Comparative sequence analysis of the mouse Mtm locus and the
RT   corresponding region of human Xq28.";
RL   Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF125313; AAF22120.1; -.
SQ   SEQUENCE  756 AA;  81963 MW;  8F5B672B378A5943 CRC64;

Query Match             5.0%;  Score 184.5;  DB 11;  Length 756;
Best Local Similarity  21.6%;  Pred. No. 0.00018;
Matches  106;  Conservative  70;  Mismatches 179;  Indels 135;  Gaps  22;

Qy  261  ATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDD 320
              : ||:|:|   |: |||:|:|  :  ::|::     :::     ||
Db   81  STADTGLKGHAVP-------YYEKNN--SMPAVDQELQDLLEELTKIQEPSSNDL---D 127

Qy  321  VVGILGQKRFWRARRAIANQQRVFA------VQLFELHRLIKVQKLIAASPDLLLDEISF 374
         :| ||||||| |  ||                |||||| || :  |:|   || : ::
Db  128  LEKIIGSKP--EEPLVLHNPQAPLGPPAKLPVQMPHMESLGSSKEFASSCSQVAGTSLPI 185

Qy  375  LGKVSAKSYPV----KKLLPS-----------EFLVKPPLPHVVVKQRGDSEKTDQHK 417
         ||| |  |          |||           :||  :  ||
Db  186  MPSSTGMSYSYIPSSSKQIVSSSSTAQAQVKNQVQNMLPVTMPPLSVPQWHHA-----HQ 240

Qy  418  MESSAENVVGRLSNQGHHQQ-SNYMPFANNPPASPAPNGYCFPPQPP----------- 464
         |:::       ||  :||||  ||  |  |  | :| ||    ||
Db  241  LKALLAASKQGSATKQGSNRNWSSLPPPGLSPPYLLPVPSPHPPPQPPPPPCSPQNFTASC 300

Qy  465  ------SGNHQQ-----WLIPVMSPSEG----------LIYKPH--PGM 490
               :|||||      | | ||| |            |||||   |||
Db  301  MSSSSLSGGAVQSSPNALLSSMAPSSNASLGPTLPYVPAKLPGLPLNQQPQFSPQSSILA 360

Qy  491  -AHTGHYGGYYGHYM-----PTPMVMPQYHPGMGFPP--PGNGYFPPYGMMPTI--MNPY 540
          | ||| | | |||     | | ||| | | | |||  |||| | || | |   |||
Db  361  NLVSSVKSPQGHLISALPTSTPGPSPPYRPENLSSPGLPQQSFTPQYSLIRSLTPTSNL 420

Qy  541  CSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDN--EPAPQQQQPTKSYPRARKSR 597
         |:||||||||                        |  |  |||| |
Db  421  LSQQQQQQQQQQ-----QQQQQQQQQQHQANSIFKPMTSSQQPKTLS----MIMQ 471

Qy  598  QGSTGSSPSGPQ--GISGSKSFRPPAAVDEDSNI-----------NNAPE 634
         |||:||:|:|                            :::
Db  472  QGLSSSSPEAPEPTFSNTKPLSHFVSEPSPQKMASMSTHSRQSSLLHYLPQATPAHAPS 531

Qy  635  QIMTTTTTTT 644
         |: |: |:|
Db  532  ATASSTATAT 541
```

RESULT 16
Q8I9J8
ID   Q8I9J8;   PRELIMINARY;   PRT;   960 AA.
AC   Q8I9J8;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   SR-related CTD associated factor 6.
GN   SCAF6 OR CG6615 OR CG6626 OR CG14058 OR CG32168.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Oregon R;
RA   Sampson N.D., Hewitt J.E.;
RT   "Functional characterization of the novel SR-related CTD associated
RT   factor, SCAF6.";
RL   Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF536543; AAN77184.1; -.
DR   FlyBase; FBgn0052168; scaf6.
DR   GO; GO:0005622; C:intracellular; IEA.
DR   GO; GO:0003723; F:RNA binding; IEA.
DR   GO; GO:0006396; P:RNA processing; IEA.
DR   InterPro; IPR000467; G_patch.
DR   InterPro; IPR000061; Surp.
DR   Pfam; PF01585; G-patch; 1.
DR   Pfam; PF01805; Surp; 1.
DR   SMART; SM00443; G_patch; 1.
DR   SMART; SM00648; SWAP; 1.
DR   PROSITE; PS50174; G_PATCH; 1.
DR   PROSITE; PS50128; SURP; 1.
SQ   SEQUENCE 960 AA; 107559 MW; CE968612D75EB72F CRC64;

```
Query Match             5.0%;  Score 183.5;  DB 5;   Length 960;
Best Local Similarity  20.3%;  Pred. No. 0.00029;
Matches  128;  Conservative  67;  Mismatches 203;  Indels 231;  Gaps  27;

Qy   78  ERNLSVQHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRS 137
             : ::| | |  |:  ||  |:     :  ::| ||::||                  ||
Db  172  EANLSAQH--TALMNQKTKQIEDAIAAAQTAQQEQMAGEQGIVLRDFDGVLQPIIESCTK 229

Qy  138  QSHGRTKSGI--------------------------------------EKEKHTPMVAPSSHHSIRFQE 168
         ||| |||||                                       ::
Db  230  DSISAGKNWILQHSTDSAKINVLLIYLLKKALVNGSTFQQKLHLIYLVNDILHHCMRKNI 289

Qy  169  VNQTGSKQNVCLAT-CS------------KPEVRDQVKANARSGGFVISLDVSVTEEID--LEKSA 219
         || ||| |  |   |              ||  ||  || | |::  |:  |:::| |   |:
Db  290  NDLKNSLENVVIPMFCSADLIGNPDQRQKLAK----LLSLWESKAKFFACVISKLQ 342

Qy  220  SSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDND 279
          :|::::   |   |:| :::|                   | :  |:| |::|  |:||
Db  343  SPDSSMQEY------------------KINLQNT-------HHDITAKFTQNTKATLDNYQKQ 380

Qy  280  REYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGIIGQKRFWRARKAIAN 339
         || :   ||  || :  : ||  ::::| |:|:  |:|   ||  ||||| ||||:|| |
Db  381  HHVFIQHASQQIVQLEKQAQH----------------------------------------L 402

Qy  340  QQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAKSYPVKKLLLPSEFLVKPPL 399
         || |: |||:|| ||:||  :||                :|   |
Db  403  EQQIVAQKSQQQHMLQHQQKNI----------PSLMSHRIA------------------M 434

Qy  400  PHVVVKQRGDSEKTDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAPNGYCFP 459
             :::|||                                        ::  |
Db  435  P-----GDHE------------------HMNSQGPHDQQH--PQAGN--NMYPGGN-FQ 465

Qy  460  PQPPPSGNH----------QQWLIPVMSPSEGLLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQY 513
          |: |||||           |||| |  ||| | |||||| |  ||  | | || |  |
Db  466  QQQPPPGNHFADPRGPNFFIPDMS----------KPPPG---------FPGPMHHNQGPAMGQQQP 512

Qy  514  HPGMGFPP----PGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQ 570
         |||:|:|     |||| | ||| |||: |  |:||||||||||  | | ||| |||:|||
Db  513  QPGQ-FPPMQAPGSNEPPVDLGVLSAAIQACLQMQQQHPDDPQQQGQH--QLQQQQQQQ 569

Qy  571  QQRSDNEPAPQQQQQ--PTKSYPR-ARKSRQGTGSSPSGPQGISGSKSFRPPAAVDED 626
                           |:|: |||  || :|:|                   |:
Db  570  Q------------PQQQKQPPPMAAYPQNIVRPGQQNADSS--------------EVD 601
```

```
Qy    627  SNINNAPEQTMTTTTTTTTRTTVTQTTRDG 655
                :.||
Db    602  LDVLNA--------AIRAVIINKPEDG 620


RESULT 17
O08778
ID   O08778        PRELIMINARY;        PRT;   3938 AA.
AC   O08778;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Bassoon.
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Sprague Dawley; TISSUE=Brain;
RX   MEDLINE=98345363; PubMed=9679147;
RA   Dieck S., Sanmarti-Vila L., Langnaese K., Richter K., Kindler S.,
RA   Soyke A., Wex H., Smalla K.H., Kampf U., Franzer J.T., Stumm M.,
RA   Garner C.C., Gundelfinger E.D.;
RT   "Bassoon, a novel zinc-finger CAG/Glutamine-repeat protein selectively
RT   localized at the active zone of presynaptic nerve terminals.";
RL   J. Cell Biol. 142:499-509(1998).
DR   EMBL; Y16563; CAA76287.1; -.
DR   PIR; T42761; T42761.
DR   GO; GO:0045202; C:synaptic junction; IEA.
DR   GO; GO:0046872; F:metal ion binding; IEA.
DR   InterPro; IPR008899; Znf_piccolo.
DR   Pfam; PF05715; Zf_piccolo; 2.
SQ   SEQUENCE 3938 AA;  418417 MW;  5BF3C230E2C71AE2 CRC64;

Query Match             5.0%;  Score 182.5;  DB 11;  Length 3938;
Best Local Similarity   21.2%;  Pred. No. 0.0025;
Matches  170;  Conservative  80;  Mismatches  315;  Indels  235;  Gaps  38;

Qy     4  GKD--EEKILEPMFPRLHVNDADKGGPRAP----PRNKMALYEQLSIPSQRFGDHGTMNSR 58
           ||| ||                       ||
Db  3261  GKDPGEPAVLE-------------------GPTLPCCYGRGEESEEDSYDPRGKSGHHRSMESN 3306

Qy    59  SNNTSTLVH------------------PGPSSQPCGVERNLSVQHLDSSAA--NQATEKFVS 100
                                      :|: :  |   :   : : : ||
Db  3307  GRPASTHYYSDSDYRHGARADKYGPGPMG-PKHPSKNLAPAALSSKRSKHRKQGMEQKIS 3365

Qy   101  QMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRTKSGIEKEKHTPMVAPSS 160
            :|::             |::|:                           :|
Db  3366  KFSPIEEAKD----------------VESDLA----------------------SYPPPTVSSS 3391

Qy   161  --HHSIRFQEVNQTGSKQNVC------------LATCSKPEVRDQV---------KANARSGGFVI 203
             |:|:             |:|:             : :               ||
Db  3392  LTSRSRKFQDEITYGLKKNVYEQRYYGVSSRDTAEEDRMYGGSSRSRVASAYSG--------3447

Qy   204  SLDVSVTEIDLEKSASSHD---RVNDYNASLRQESRNRLYRDG---------GKTR------ 248
                                 |:| :|      :                   ||
Db  3448  -------------EKLSSHDFSSRSKGYER--ERETAQRLQKAGPKPSSLSMAHGRARPP 3492

Qy   249  LKDTDNGASHLATENHSQEGHGSPEDINDREYSKSRACASLQQINEEASDD--------- 301
                                    :|:     :     :  :       ||
Db  3493  MRSQASEEESPVSPLGRPRPAGGALPPGDTCPQFCSSH--SMPDVQEHVKDGPRAHAYK 3549

Qy   302  ----VSDDSMVDSISSIDVSPPDVVGILGOKRF-WRARKAIANQQRVFAVQLFELHRLI 355
               : :                   |:        :||||
Db  3550  REEGYILDDSHC------VVSDSEAYHLGQEETDWFDKPRDARSDR------FRHHGGH 3596

Qy   356  KVQKLIAASPDLLLDEISFLGKVSAKSY---PVKKLLPSEFIVKPPLPHVVVKQRGDSEK 412
                                        :|                |:
Db  3597  TVSSSQKRGP-------ARHSYHDYDEPPEEGLWPHD---EGGPGRHTSAKEHRHHGD 3644

Qy   413  TDQHKMESSAENVVGRLSNQGHHQQSNYMPFANNPPASPAP------------NGYCFPPQPPPS 465
```

```
Db  3645  HGRHSGRHAGE--PGRRAARPHARDMGRHETRPHPQASPAPAMOKKGQPGY-----PSS 3697
            :  :                |:  |:           ||:           | ||||
Qy   466  GNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNG 525
            :  : : : :                                    :  ::||: :
Db  3698  ADYSQ---PSRAPSAYHHADSKKGSRQAHSGPTVLQPKPBAQAQPQMGRQAVPGPQQS 3754

Qy   526  YFPPYGMMP--TIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQRSDNEPAPQQ- 582
            ||  :   |    : :     || :|| :       |     || :  ||||||
Db  3755  QPPSSRQTPSGTASRQPQTQQQQQQQQQQQQ----QQQQQQQGLGQQAPQQA 3807

Qy   583  -------QQQ--------PTKSYPRARKSRQST---GSSPSGPQGISGSKSFRPFAAVD 624
                                : :|                :::|
Db  3808  PSQARLQQQSQPPTTRSTAPAASHP-AGKPQPGPTTAPGPQPAGLPRAEQAGSSKPAAKAP 3866

Qy   625  EDSNINNAPEQTMTTTTTTTTRTTVTQTTRDGG--------GVTRVIKVVPHHNAKLA 672
            :  ::                                       |::
Db  3867  QQ---GRAPQAQSAPGPAGAKT-----GARPGGTPGAPAGQPAAEGESVFSKILPGG---A 3916

Qy   673  SENAARIFQSIQEERKRYDS 692
            :.  |:  ::|: :|||
Db  3917  AEQAGKLTEAVSAFGKKFSS 3936


RESULT 18
Q8IMG4
ID   Q8IMG4        PRELIMINARY;        PRT;   671 AA.
AC   Q8IMG4;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG1775-PD.
GN   MED OR CG1775.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20195006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayrahtaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E., Sun X.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
```

```
Qy    378  VSAKSYPVKLLPSEFLVKPPLP-HVVVKQRGDSEKTDQHKMESSAENVVGRLSNQG---  433
               |:|                   |  ||| :| ||  |    |
Db    130  VNPYHY------------ERVVSPGIDLSGLSLQSGPFSRLV--KDEYSAGPLVGSMDIDGNDI  178

Qy    434  ---HHQQSNYMPFANNPPASPAPNGYCPPQPPPSGNHQQWLIPVMSPSEGLLIYKPH--  487
                                                  ::|         |
Db    179  GTIQHH----------------PTQMVGPGGYGYP-------------QGPSEYAV--PHGM  209

Qy    488  PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCCSSQQQQ  547
               |||||             ||||  ||       |          |       ||||
Db    210  PGMP----------GPMNFGPVMAPPPPDQAQNPQGNGVHHTQANSPTDPASALAMQQQQ  260

Qy    548  QQQPNEQMNQFGHPGNLQNTQQQQQRSD-------------------NEPAPQQQQQ---  585
               ||||            |||||                      ::|:
Db    261  QQQQQQQQQQ-----------QQQQQSGGVPNGSVNAGGAAAGGQYYGQPPPVSQMQGA  310

Qy    586  ---PTKSYPRARK----------SRQSTGSSPSGPQGISGKSFRPFAAVDEDSNINNAPEQT  636
                  ||:  |||
Db    311  GGGGTSVAPSVHAQQNGYVSQPGSAGSAPVGGGGVFGTAQPTP-----------QQPQQP  359

Qy    637  MTTTTTTRTTVTQTTRDGG  656
               ||
Db    360  PTCVQANTGSAGAQAGAGGG  379
```

RESULT 19
Q8IMG6
```
ID   Q8IMG6                  PRELIMINARY;    PRT;    697 AA.
AC   Q8IMG6;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG1775-PB.
GN   MED OR CG1775.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.H., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA   Abril J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spradling A.C., Stapleton M., Strong R., Sun E.,
```

```
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003779; AAN14279.1; -.
DR   FlyBase; FBgn0011655; Med.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IGI.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti.. ..; NAS.
DR   GO; GO:0008283; P:cell proliferation; IMP.
DR   GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR   GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
DR   GO; GO:0007476; P:wing morphogenesis; IMP.
DR   InterPro; IPR001132; Dwarfin.
DR   InterPro; IPR003619; Dwarfin_A.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF03165; MH1; 1.
DR   Pfam; PF03166; MH2; 1.
DR   SMART; SM00523; DWA; 1.
DR   SMART; SM00524; DWB; 1.
SQ   SEQUENCE   671 AA;   71127 MW;   3B5F373690F434FC CRC64;
```

```
Query Match             5.0%;   Score 182;   DB 5;   Length 671;
Best Local Similarity  24.3%;   Pred. No. 0.00022;
  Matches  107;   Conservative   42;   Mismatches 155;   Indels  136;   Gaps   21;

Qy    265  HSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI  324
               | |||: | :||        :||:|||    |:||:|| |||  :||||
Db     28  HRQCG------ESEGFAKRAIESLVKKLKEKRDEL--DSLITAITTNGAHPSKCVTI  76

Qy    325  LGQKRFWRARKAIANQQ---RVFAVQLF--ELHR-LIKVQKLIAASPDLLDEISFLGK  377
               |||| | |: |             |:|    | :||  :  || | :||:||
Db     77  ---QRTLDGRLQVAGRKGFPHVIYARIWRWPDLHKNELKHVKYCAFAFDLKCDSVC----  129
```

```
Db  103  ---QRTLDGRLQVAGRKGFPHV1YAR1WRWPDLHKNELKHVKYCAFAFDLKCDSVC----  155
              :|::    : :|| : ||::  :    ||| :|| ||:|:||   |:|
Qy  378  VSAKSYPVKKLLPSBFLVKPPLP-HVVVKQRGDSEKTDQHKMESSAENVVGRLSNOG---  433

Db  156  VNPYHY-----ERVVSPGIDLSLSLQSGPSRLV--KDEYSAGPLVGSMDIDGNDI       204
         :|:        | : :| | |: |  : |||  :  |:: : | | : :| :|
Qy  434  ---HHQQSNYMPFANNPASPAPNGYCFPPQPPPSGNHQQWLIPVMSPSGLIYKPH--     487

Db  205  GTIQHH------------PTQMVGPGGYGYP--                             235
         | :|::              ||  |||||||
Qy  488  PGMAHTGHYGGYYGHYPTPMVMPQYHPGMGFPPPGNGYFPPYGMMPTIMNPYCSSQQQQ   547

Db  236  PGMP---------GPMNPGPVMAPPPPQQAQNPQGNGVHHTQANSPTDPASALAMQQQQQ   286
         |||          ||  | | :|||||  :|  : | ||                :||||
Qy  548  QQQPNEQMNQFGHPGNLQNTQQQQRSD------------NEPAPQQQQQ---          585

Db  287  QQQQQQQQQQ---QQQQQSGGVPNGSVNAGGGAAAGGQYYGQPPPVSQMQGA          336
         ||||||||||   |||||
Qy  586  ---PTKSYPRARK-------SRCGSTGSSPSGPQGISGSKSFRPFAAVDEDSNINNAPEQT 636

Db  337  GGGGTSVAPSVHAQQNGYVSQPGSAGSAPVGGGGVFGTAQPTP--------QPQQP      385
                                                                :
Qy  637  MTTTTTTTRTTVTQTTRDGG                                          656
         |||
Db  386  PTGVQANTGSAGAQAGAGGG                                          405
```

RESULT 20
Q9V6W7

ID   Q9V6W7   PRELIMINARY;   PRT;   1594 AA.
AC   Q9V6W7; Q9V6W8;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   CG8118 protein.
GN   MAM OR CG8118.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Murray J.M., Nelson D.L.,
RA   Nelson S.M., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., WoodageT., Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA   Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhang J., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Puri V., Richards S., Scheeler F.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RP   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003779; AAN14277.1; -.
DR   FlyBase; FBgn0011655; Med.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IGI.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti...; NAS.
DR   GO; GO:0008283; P:cell proliferation; IMP.
DR   GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR   GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
DR   GO; GO:0007476; P:wing morphogenesis; IMP.
DR   InterPro; IPR001132; Dwarfin.
DR   InterPro; IPR003619; Dwarfin_A.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF03165; MH1; 1.
DR   Pfam; PF03166; MH2; 1.
DR   SMART; SM00523; DWA; 1.
DR   SMART; SM00524; DWB; 1.
SQ   SEQUENCE   697 AA; 73679 MW; 222398AA59F1D28E CRC64;

```
Query Match             5.0%;        Score 182; DB 5; Length 697;
Best Local Similarity   24.3%;       Pred. No. 0.00024;
Matches   107;  Conservative   42;  Mismatches  155;  Indels  136;  Gaps  21;

Qy  265  HSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI  324
              ::      : :||  :|| :   |::| |:     :  :|| :|  |  ||| ::|
Db   54  HRQGG-------BSEGFAKRAIBSLVKKLKEKRDEL--DSLITAITTNGAHPSKCCVTI   102

Qy  325  LGQKRFWRARKAIANCQ--RVFAVQLF--ELHR-LIKVQKLIAASPDLLLDEISFLGK    377
```

```
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AE003816; AAF58300.2; -.
DR   FlyBase; FBgn0002643; mam.
DR   GO; GO:0005634; C:nucleus; IDA.
DR   GO; GO:0007500; P:mesoderm cell fate determination; IMP.
DR   GO; GO:0007399; P:neurogenesis; NAS.
SQ   SEQUENCE   1594 AA;  167262 MW;  D4236DA26F70D092 CRC64.
```

```
Query Match             4.9%;  Score 180;  DB 5;  Length 1594;
Best Local Similarity   18.2%;  Pred. No. 0.0011;
Matches  162;  Conservative  95;  Mismatches 306;  Indels 328;  Gaps  35;

Qy      40 EQLSIPSQRFGDHGTMNSRSNNTSTLVHPGPSSQPCGVERNLSVQHLDSSAANQATEKFV 99
              ::  :  :  :  :::  :: :  ::  :  :::
Db     166 QETTVLQKRFLE--SKNKRAAKKTDKKLPDPSQOHQQQOQQHQQAQAQTMLA 223

Qy     100 SQMSFMENVRSSAQHDQRKMVREEED-------FAVPVYINSRRSQSHGRTKSGIEKBKH 152
                 :    ::   :  ::   :     ::  :
Db     224 GQL-----QSSVHVQQKFLKRPAEDVDNGPDSFEPPHKLPNNNNSNSNNNGNANANN 277

Qy     153 TPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTBE 212
                :         :    ::  :  :   :      :::
Db     278 ------GGNGSNTGNNTNNNGNSTN------YNASLRQE-----NNGGSNNNGSENLTKFSVEIVQQ 319

Qy     213 IDLEKSA--SSHDRVND------YNASLRQE 247
            ::  :       :    ::     :::
Db     320 LEFTTSAANSQPQQISTNVTVKALTNTSVKSEPGVGGGGGGGSGGGGGGGSGNNNNNNNGGGGGGG 379
```

```
Qy     248 RLKDTDNGAESHLATENHSQE----------------------------GHGS 272
             :  :: :: :  : :
Db     273 PEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSLSSIDVSPDVVGILGQKRFWR 332

Qy     273 PEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSLSSIDVSPDVVGILGQKRFWR 332
             ::  ::  ::  :: :::
Db     333 NGNNNNNGGDHH--QQQHQQQQQQGGLGLGNNGRGGGPGMATGPGGVAGGLGMGM 437

Qy     380 NGNNNNNGGDHH--QQQHQQQQQQGGLGLGNNGRGGGPGMATGPGGVAGGLGMGM 437
             : : :
Db     438 PPNMMSAQQKSALGNLANLIVECKREPDHDFPD-----LGSLDKD-----GGGGQ----- 481

Qy     333 ARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLDE----------------IS 373
              :  :  ::   ::  :: :  ::  ::  ::  :
Db     482 ---------------FPGFPDLLGDDNSENNDTFKDLINNLQDFNPS 513

Qy     374 FLGKVSAKSYPVKKLLPSEFL-VKPP------------------LPHVVVK 405
              :  :     :  :  :: :     : :
Db     514 FLDGFDEK--PLLDIKTEDGIKVEPPNAQDLINSLNVKSEGGLGHGFGGFGLGLDNPGMK 571

Qy     406 QRGDSE--------KTDQHKMESSAENVVGRLSNOGH------------HQQSNYM 441
              :  :::   :  :  :  :  ::   :  ::
Db     572 MRGGNPGNQGGFPNGPNGGTGGAPNAGGNGGNSGNLMSEHPLAAQTLKQMAEQHQHKNAM 631

Qy     442 PFANNPPASPAPNGYCFPPQPPPSGNHQQWLLPVMSPSEGLIYKPHPGMAHTGHYGGY--- 499
              :  :  :::   ::   ::  :::  ::::
Db     632 --------GGMGGFPRPHGMNFPQQQQQQQQQQQAQQQHGQMMGGQPGRNID 679

Qy     500 YGHYMPTPM----------------VMPQYHPGMGFPPPNGNYFPPYPGMMPTIM 537
              :  :  :  :   :  ::   : :
Db     680 YGGGFPNDFGLGPNGPQQQQAQQQPQQQHLPPQFHQQKG--PGPGAGMNVQQNFLDIKQ 738

Qy     538 NPYCSS----------------QQQQQQQPNEQMNQ---------FGHPGNLQNTQQQQQ--- 572
              :    :::  :::  :  :  :
Db     739 ELFYSSQNDFDLKRLQQQAMQQQQQQHHQQQQQQPKMGGVPNFNKQQQQQQVPQQQ 798

Qy     573 -RSDNEPFAPQQQQQQPTKSY------PRARKS------RQGSTGSSP----------S 606
               : :  :  ::   :::  :
Db     799 LQQQQQQQQQQQQQYSPFSNQNPNAAANFLNCPPRGGPNGNQOPGNLAQQQQQPGA 858

Qy     607 GPQGISGSKSFRPFAAVDEDSNIN-----------NAPEQTMTTTTT--RTTVTQTTRD 654
             ::  :   :  :::  :  :
Db     859 GPQ---QQQRGNAANGQQNNPNAGPGGNTPNAPQQQQQQSTTTLQMKQTQQLHISQQ 914

Qy     655 GGGVTRV-IKVVPH---------NAKLASENAARIFQSIQEERKRYDSSKP 695
              :::   :  ::    :::   :
Db     915 GGGAHGIQVSAGQHLHLSGDMKSNVSVAAQQG--VFFSQQAQQQQQQQQP 963
```

```
RESULT 21
Q86IN4
ID   Q86IN4                PRELIMINARY;       PRT;  1749 AA.
AC   Q86IN4;
DT   01-JUN-2003 (TrEMBLrel. 24, Created)
DT   01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Similar to exonuclease ii.
OS   Dictyostelium discoideum (slime mold).
OC   Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
OX   NCBI_TaxID=44689;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RX   MEDLINE=22092622; PubMed=12097910;
RA   Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,
RA   Lehmann R., Baumgart C., Parra G., Abril J.F., Guigo R., Kumpf K.,
RA   Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.;
RT   "Sequence and analysis of chromosome 2 of Dictyostelium discoideum.";
RL   Nature 418:79-85(2002).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=AX4;
RA   Baumgart C.;
RL   Submitted (MAR-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AC116908; AAO52009.1; -.
DR   GO; GO:0005634; C:nucleus; IEA.
```

```
DR   GO; GO:0004527; F:exonuclease activity; IEA.
DR   InterPro; IPR004859; Put_53exo.
DR   Pfam; PF03159; XRN_N; 1.
KW   Exonuclease.
SQ   SEQUENCE   1749 AA;  200750 MW;  991F704A3592B2D6 CRC64;

Query Match             4.9%;
Best Local Similarity   18.1%;  Pred. No. 0.0012;
Matches  141;  Conservative 117;  Mismatches 245;  Indels 278;  Gaps 36;

Qy      6 DEEKILEPMPPRLHVNDADKG--GPRAPPRN---------------------KMALYEQLSIPS 46
            :::||::  |: || |: :: | |:::                         :::|| :|
Db   1111 ERHHLIKGLQPKQDLNNSQNGVKSPTLSSQNYSFQSFANTGKFHLGDRVVSILDKGNLP- 1169

Qy     47 QRFGDHGTMNSRSNNTSTLV-----HPGPSSQP-CGVERNLSVQHL---------------DSS 89
           ||| ||  : ::                :|::::                            |::
Db   1170 --FGTFGTVASIQDGKVDVVFDTECFAGNSLDGYCSEKRGICISKLRLYNLSCPPPPPKS 1227

Qy     90 AANQATEKFVSQMSFMENVR----SSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRTKS 145
            :|:  ::: :: |::                          ::| |||:    :|::|
Db   1228 TINKFYDQSIDPAEYWEKVQTQQNNNHHGQKKI------------YSNGGQKL 1269

Qy    146 GIEKEKHTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISL 205
            |:::|| ::                   |::                      |:|:
Db   1270 NQQIDQQTPITNAVENKELNWQQLQ-----LINNISNPQ 1303

Qy    206 DVSVTEEIDLEKSASSHDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATENH 265
            :|:| |||  :                        |:            || :|:||::
Db   1304 ------QHNANNMNNNNYN------------------------------NNNNNHHGQNH 1329

Qy    266 SQ-EGHGSPEDIDNDREYS--KSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDV 321
            || :|||::|:|| |||:    ::|    |||:|::::|:||:::  :::|:  ::|:||
Db   1330 NQNKVNQHPLAINNPNSVNYPMKRKPTYVKQNFEQQEYADLS-----KNYPNLEYN--- 1380

Qy    322 VGILGQKFWRARKAIANQQRVFAVQLFELHRLIKVQKLIAASPDLLLDEISFLGKVSAK 381
            |||||||| :  :  :::     ::              :::    :|     :|::|
Db   1381 -------FYYDQ-----NDQRKQPQQL---QQPKPQQQPQPQPQ-------------PQP 1412

Qy    382 SYPVKKLLPSEFLVKPPLPHVVKQRGDSEKTDQHKMESSAENVV--------GRLSN 431
            ::                   :|::                             ||:|
Db   1413 KQPKQPKQPKQSKQPQPQQPQQPQQEPIDPEKLRQQTRTNKRLNLIYQNIEQNSFPGSQSS 1472

Qy    432 QGHH---------QQSNYMPFA----------------NNP--PASPAPN--GYCFPPQP 462
            ||:           ||||||||                |: :  :     |: ||
Db   1473 EHNNGDGSSEEQVQTNPNALLLNNNMFASTSISSDQTNDPDDGLPQGPPPHMMGH-YPPGP 1531

Qy    463 PPSGNHQQWLIPVMSPSEGLIYKPHPGMAHTGHYGGYYCHYMPTPMVMPQYHPG------- 516
            ||                                   :|: :
Db   1532 PPM----------------MGYPPH---YHPGH 1567

Qy    517 MGFPP---------PGNGYFPPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQ- 568
            |||: :                :    |||:  || ||||||:|  :|| |  ||::|:
Db   1568 MGYPPHYHPGHPYPPHHPGQQEEHHHHQQQQEQQOQHPTQQEEPQNQHPKKKQPKQPKLPK 1627

Qy    569 ------QQQQ----------RSDNEPA-PQQQQQPTK---SYPR-ARKSRQ 598
                  ||||                 :|::                  :| : |||::|
Db   1628 QPNQNQTQPTQDGQQQQPQPKQPKQPNQTPKQPAQPKQPKQPAQPKQPAQPKQAAQTKQ 1687

Qy    599 GSTGSSPSGPQGISGSK------SFRPFAAVDEDSNINNAPEQTMTTTTTTRTTVTQT 651
            |:|::    |  ::|||      ::|||:|:|| |:: :  ||:|::|::::|::||:|
Db   1688 PAQPKQPAQPKQAAQTKQPKQPAQPKQPKQPNQTPKQP-AQQTSPTTPNPTTETTSTPPTPT 1746

Qy    652 T 652
          |
Db   1747 T 1747
```

```
RESULT 22
Q96JK9
ID   Q96JK9     PRELIMINARY;        PRT;  1162 AA.
AC   Q96JK9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
```

```
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical protein KIAA1816 (Fragment).
GN   KIAA1816.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain;
RX   MEDLINE=21245130; PubMed=11347906;
RA   Nagase T., Nakayama M., Nakajima D., Kikuno R., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XX.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 8:85-95(2001).
DR   EMBL; AB058719; BAB47445.1; -.
DR   Genew; HGNC:16272; MAML3.
KW   Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   1162 AA;  124759 MW;  0447AF72DC0B877B CRC64;

Query Match             4.9%;   Score 179.5;  DB 4;  Length 1162;
Best Local Similarity  19.0%;   Pred. No. 0.00073;
Matches  135;  Conservative  91;  Mismatches 232;  Indels 253;  Gaps  30;

Qy    125 DFAVP----------------------VYINSRRSQSHGRTKSGIEKEKHTPMVAPSSHH- 162
            ||::|                       |:  ::|:  :::||:|:|::|:|||  ||||
Db     32 DFAAPAAAAANGSSICINSSLNSSLGGAGIGVNNTPNSTPAAPSSNHPAAGGCGGSGGPGG 91

Qy    163 -SIRFQEVNQTGSKQNVCLATCSKPEVRDQVKANARSGGFVISLDVSVTEEIDLEKSASS 221
             |:::                                             :::|:||
Db     92 GSAAVPKHSTVVERLRQRIEGCRRHHVNCENR-----------------YQQAQVEQLELER 136

Qy    222 HDRVNDYNASLRQESRNRLYRDGGKTRLKDTDNGAESHLATE-NHS--------------QEG 269
            | |:  :::|:::                                  :||:
Db    137 RDTVSLYQRTLEQRAKKSGAGTGKQQHPSKPQQDAEAASAEQRNHTLIMLQETVKRKLEG 196

Qy    270 HGSPEDIDNDREYSKSRAC-ASLQQINEEASDDVSDDSMVDSISSI---------DVSPDDV 321
            |                   ::  |:|:|::::|                        ||
Db    197 ARSP--LNGDQQ---NGACDGNFSPTSKRIRKDIS--AGMEAINNLPSNMPLPSASP-- 246

Qy    322 VGILGQKFWRARKAIANQQRVFAVQLFELHRL----------IKVQKLIAASPDLLLDEISFLG 376
            |||||||||:| :   ::                        |:   |:  :   : : : ::|
Db    247 ----------------LHQLDLKPSLPLQNSGTHTPG-LLEDLSKNG 276

Qy    377 KVSAKSYPVK---------------KLLPSEFLVKPPLPHVVKQRGDSEKTDQHKMESS---- 421
            ||                       |::  |||| |::||: ::
Db    277 RLPEIKLPVNGCSDLEDSFTILQSKDLKQEPLDDPTCIDTSETSLSNQNKLFSDINLNDQ 336

Qy    422 ------------------AENVVGRLSNQGHHQQS 444
                              :|:::| ::|::|||
Db    337 EWQELIDELANTVPEDDIQDLFNEDFEEKKEPEFSQPATETPLSQESASVKSDPSHSPFA 396

Qy    445 N---NPPASPAPNGYCFPP------------------QPPPSGNHQQWLIPVMSPSEGLIYKP 486
             :  |||:|:|||::||                    || :|| |||:
Db    397 HVSMGSPQARPSSSG---PPFSTVSTATSLPSVASTPAAPN-----PASSPANCAVQSP 447

Qy    487 H-PGMAHTGHYGGYYCHYMPTPMVMPQYHPGMGFPPPGNGYF------------------- 527
            | |||:||||||||:|||||||::|||:|||:|||||| :|
Db    448 QTPNQAHT---------------------PGQAPPRPGNGYLLNPAAVTVAGSASGPVAV 486

Qy    528 PPYGMMPTIMNPYCSSQQQQQQQPNEQMNQFGHPGNLQNTQQQQQRSDNEPAPQQQQQPT 587
            || :|:        ||||||||:   |:  |  ||:|:|:||::||:
Db    487 PSSDMSPAEQLKQMAAQQQQRAKLMQQKQQ------------QQQQQQQ 531

Qy    588 KSYPRARKSRQGSTGSSPSGPQG--ISGSKSFRPFAAVDEDSNINN-APEQTMTTTTTTT 644
            | :|:       :| |::|  |    :||||:  :||::|:||:|| ||: |:|::|:||
Db    532 QQQQHSNQTSNWSPLGPSSPYGAAFTAEKPNSPMMYPQAFNNQNPIVPFMANNLQKTTM 591

Qy    645 RTTVTQTTRDGGGVTRVIKVVPHNAKLLASENAARIFQSIQEERKRYDSSKP 695
            ::|:|:     :|:|:::                      : ||  |:|:|
Db    592 NNYLPQNHMN-----MINQQPNNLGTNSLNK------QHNLITYGNTKP 629
```

```
RESULT 23
O77069
ID   O77069      PRELIMINARY;      PRT;   697 AA.
AC   O77069;
DT   01-NOV-1998 (TrEMBLrel. 08, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   MEDEA-B.
GN   MED OR MEDEA OR CG1775.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=98361878; PubMed=9694800;
RA   Xu X., Yin Z., Hudson J.B., Ferguson E.L., Frasch M.;
RT   "Smad proteins act in combination with synergistic and antagonistic
RT   regulators to target Dpp responses to the Drosophila mesoderm.";
RL   Genes Dev. 12:2354-2370(1998).
DR   EMBL; AF019753; AAC35436.1; -.
DR   HSSP; Q13485; 1DD1.
DR   TRANSFAC; T04382; -.
DR   FlyBase; FBgn0011655; Med.
DR   GO; GO:0005737; C:cytoplasm; IDA.
DR   GO; GO:0005634; C:nucleus; IGI.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti....; NAS.
DR   GO; GO:0008283; P:cell proliferation; IMP.
DR   GO; GO:0009950; P:dorsal/ventral axis specification; IMP.
DR   GO; GO:0007480; P:leg morphogenesis (sensu Holometabola); IMP.
DR   GO; GO:0006355; P:regulation of transcription, DNA-dependent; NAS.
DR   GO; GO:0007179; P:TGFbeta receptor signaling pathway; IGI.
DR   GO; GO:0007476; P:wing morphogenesis; IMP.
DR   InterPro; IPR001132; Dwarfin.
DR   InterPro; IPR003619; Dwarfin_A.
DR   InterPro; IPR008984; SMAD_FHA.
DR   Pfam; PF03165; MH1; 1.
DR   Pfam; PF03166; MH2; 1.
DR   SMART; SM00523; DWA; 1.
DR   SMART; SM00524; DWB; 1.
SQ   SEQUENCE   697 AA;  73668 MW;  9C33C49E1B3ACC5C CRC64;

Query Match             4.9%;  Score 178.5;  DB 5;  Length 697;
Best Local Similarity  23.1%;  Pred. No. 0.00042;
Matches  116;  Conservative  50;  Mismatches 177;  Indels 159;  Gaps  23;

Qy    265 HSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDVSDDSMVDSISSIDVSPDDVVGI 324
Db     54 HRQGG-----------ESEGFAKRAIESLVKKLKEKRDEL--DSLITAITTNGAHPSKCVTI 102

Qy    325 LGQKRFWRARKAIANQQ---RVFAVQLF--ELHR-LIKVQKLIAASPDLLDEISFLGK 377
Db    103 --QRTLDGRLQVAGRKGKFPHVIYARIWRWPDLHKNELKHVKYCAFAFDLKCDSVC---- 155

Qy    378 VSAKSYPVKKLLPSEFLVKPPLP-HVVVKQRGDSEKTDQHKMESSAENVVGRLSNQG-- 433
Db    156 VNPTHY------ERVVSPGIDLSGLSLQSGPSRLIV--KDEYSAGPLVGSMDIDGNDI 204

Qy    434 ---HHQQSNYMPFANNPPASPAPNGYCFPPPQPPPSGNHQQWLIPVMSPSEGLIYKPH-- 487
Db    205 GTIQHH-------------PTQMVGPGGYGYP----------------QGPSEYAV--PHGM 235

Qy    488 PGMAHTGHYGGYYGHYMPTPMVMPQYHPGMGFPPPGNGYFPPYGMPTIMNPYCSSQQQQ 547
Db    236 PGMP-------------GPMNPGPVMAPPPPQQAQNPQGNGVHHTQANSPTDPASALAMQQQQ 286

Qy    548 QQPNEQMNQFGHPGNLQNTQQQQQQRSD 585
Db    287 QQQQQQQQQ----------------QQQQQQSGGVPNGSVNAGGGAAAGGQYYGQQPPVSQMQGA 336
```

```
          Qy    586 ---PTKSYPRARK-------SRQGSTGSSPSGPQQISGSKSFRPFAAVDEDSNINNAPEQT 636
                    ::: :::   :|::  :::   :::  :::::  :::
          Db    337 GGGGTSVAPSVHAQQNGYVSQPGSAGSAPVGGGGVFGTA-----------QPTSQQPQQP 385

          Qy    637 MTTTTTTTTRRTTVTQTTRDGGG-------VTRVIKVVPHNAK-----LAS 673
                     ::|:|  ::: ::      :  |
          Db    386 PTGVQANTGSAGAQAGAGGGAAGTWTGPNTLTYTQSMQPPNPRSLPGGFWNSSLSGDLGS 445

          Qy    674 ENAARIFQSIQEERKRYDSSKP 695
                    :::::    ::::::|
          Db    446 PQQTPPQQQQQQQPRLLSRQP 467
```

```
RESULT 24
Q93YP4
ID   Q93YP4      PRELIMINARY;      PRT;   1024 AA.
AC   Q93YP4;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Epsin-like protein (At3g59290).
OS   Arabidopsis thaliana (Mouse-ear cress).
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   eurosids II; Brassicales; Brassicaceae; Arabidopsis.
OX   NCBI_TaxID=3702;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA   Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA   Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA   Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA   Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA   Ecker J., Theologis A., Davis R.W.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RA   Cheuk R., Chen H., Kim C.J., Shinn P., Bowser L., Carninci P.,
RA   Dale J.M., Hayashizaki Y., Hsuan V.W., Ishida J., Jones T., Kamiya A.,
RA   Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA   Nguyen M., Onodera C.A., Palm C.J., Quach H.L., Sakurai T., Satou M.,
RA   Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA   Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT   "Arabidopsis ORF clones.";
RL   Submitted (MAY-2003) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY059878; AAL24360.1; -.
DR   EMBL; BT008432; AAP37791.1; -.
DR   InterPro; IPR001026; ENTH.
DR   InterPro; IPR008942; ENTH_VHS.
DR   Pfam; PF01417; ENTH; 1.
DR   SMART; SM00273; ENTH; 1.
SQ   SEQUENCE   1024 AA;  109694 MW;  1C8A5F33EB12ED04 CRC64;

Query Match             4.9%;  Score 178.5;  DB 10;  Length 1024;
Best Local Similarity  19.2%;  Pred. No. 0.00072;
Matches  151;  Conservative  77;  Mismatches 280;  Indels 277;  Gaps  33;

Qy     84 QHLDSSAANQATEKFVSQMSFMENVRSSAQHDQRKMVREEEDFAVPVYINSRRSQSHGRT 143
Db    116 QYIDSSGKDQGS---------------NVRKKAQ--------SLVALVNDKERITEVRE 151

Qy    144 KSGIEKEK-HTPMVAPSSHHSIRFQEVNQTGSKQNVCLATCSKPE--VRDQVKANARSGG 200
Db    152 KAAANRDKYHNSMHRPSGGYGDKYDYEGRYGDREGRSSYGKEREYGYRDDR-NSRDGD 210

Qy    201 FVISLLDVSVTEEIIDLEKSASSHDR------------------VNDYNASL-RQESRNRLYR 242
Db    211 ----------------------------RYSRRDSEDRYGRDGNTDDEYRGRSRSVDNYNGSRGRSSDRERPIE 255

Qy    243 DGGKTRLKDTDNGAESHLATENHSQEGHGSPEDIDNDREYSKSRACASLQQINEEASDDV 302
Db    256 DDGQS--SSRDSGAPA-----DDHSQDDGRGGLE-----RKFSEQNIGAAPSYEEAVSESR 304
```

```
Qy    303  SD-DSMVDSISSIDVSPDDVVGIIGQKRFWRARKA--IANQQRVFAVQLFELHRLLIKVQK  359
                  ::  ::|  |::|||             |  ::|:
Db    305  SPVYSERDGGETPQVAPPGAAASPLAENISVDNKAADFVNESSPQQVEAF----------  354

Qy    360  LIRASPDLLLDEISFLGKVSAKSYP-----VKKLLPSEFLVKP---------------  397
                  ||:|||            :|:||
Db    355  --------DEFDPRGSVSAACAPTAGASVPAPIPPTVVSTPAPPASINAEMDLLGSLS  404

Qy    398  ---PLPHVVVKQRGDSEKTDQH--------KMESSAE----------NVVGRLSNQ  432
              ::|                   ||: |                 :|:
Db    405  DVFSPNPLAIVTSDSTSVETNGQANTGLAPSFSTSQSSTQPFDDFPGDSPFKAITSADTE  464

Qy    433  GHHQQSNYMPFANNPPAS-------------------------PAPNGYCFP-------  459
                       :|||                               || ||
Db    465  TSQHQSFGVPFQPTPPTSNPNNEHFGFGEAFSAVTDSEPGVQNMQAPPNLSVFPQEQFD  524

Qy    460  --------------------PQPP-------PSGN-----HQ------QWLI  473
                               :|||       :||     ||      ||:
Db    525  TSQSEIDILAGILPPSGPPVSLSPQDSTMPTSQFHPNGNSYESYHHQAAPTDLNMQGQT  584

Qy    474  PVMBPSEGLIYKPHPGMAHTG---HYGGYY--GHYMPTMVMPQYHPGMGFPPPGNGYF  527
                 ||                      |:  :|||    |:::||||
Db    585  PFGQASQQFNMVSHSQNHHEGMQFNNGGFTQQPGYAGPATSQPPQYTPGVSSHPPSES-F  643

Qy    528  P--------------------PYGMMPTI----------MNPYCSSQQQQ------  555
                                    |||||||           |:::| |||
Db    644  PHQPGSATSASSASSQTPYATTPNVSAGQFDGGSFMTQQPYGVTQQVHVVPSHIPQRTQSGPV  703

Qy    556  NQFGHPGNLQNTQQQ----QQRSDNEPAPQQQQQQPTKSYPRARKSRQGSTGSSP----SGP  608
              :|                |||                |:|
Db    704  AAFGNNNNIVGDMHQPGSTPSSSSQTPYPTTPNAPSGQFDGGNFMTQQPYGVIPQVHGVP  763

Qy    609  QGISGSKSFRFPFAAVDEDSNINN--APEQIMTTTTTTRTTVTQTTRDGGGVTRVIKVV  665
               |::|                    ||::|  ||:|:
Db    764  SHIPQRTQSGPVAAHGNSNNVVGDMFSPAGLSSLETSASQPSLTP------LTGAIEIV  816

Qy    666  PHNAK  670
              |||
Db    817  PQNQK  821
```

RESULT 25

```
001505
ID   001505      PRELIMINARY;      PRT;   1239 AA.
AC   001505;
DT   01-JUL-1997 (TrEMBLrel. 04, Created)
DT   01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Hypothetical protein.
GN   C37A2.2.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RX   MEDLINE=99069613; PubMed=9851916;
RA   None;
RT   "Genome sequence of the nematode C. elegans: a platform for
RT   investigating biology. The C. elegans Sequencing Consortium.";
RL   Science 282:2012-2018(1998).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Le T.T., Kemp K., Scheet P.;
RT   "The sequence of C. elegans cosmid C37A2.";
RL   Submitted (APR-1997) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Bristol N2;
RA   Waterston R.;
RT   "Direct Submission.";
```

```
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; U97194; AAB52447.2; -.
DR   PIR; T30160; T30160.
DR   WormPep; C37A2.2; CE27747.
KW   Hypothetical protein.
SQ   SEQUENCE   1239 AA;  135436 MW;  81F6C299FDC77D85 CRC64;
```

```
Query Match             4.9%;  Score 178.5;  DB 5;  Length 1239;
Best Local Similarity   23.4%;  Pred. No. 0.00095;
Matches  137;  Conservative  66;  Mismatches 235;  Indels 147;  Gaps 32;
```

```
Qy    150  EKHTPMVAPSHHSIRFQ-EVNQTGSKQNV-CLATCS-----KPEVRDQVKANARSCGFV  202
                |||:|                 ::                 |
Db     20  EKVGDIVRPSSATALLSSLSTSSSGIEASTSCASTPTTANTLPPLIPESSVEDETADGVG  79

Qy    203  ISLDVSVTEEIDLEKSASSHDRV-----NDYNASLRQESRNRLYRDGGKTRLKDTDNGAE  257
                   |:                         |::   |:        ||:
Db     80  IATE-KLVKSLIVPSPSDTARVTSAISNDTVASGKTNEEDR-----KSIDTSVD  128

Qy    258  SHLATENHSQEGHGSPEDIDN--DREYSKSRACASLQ--QINEEASDDVSDDSMVDSIS  312
                |                    :          :|||
Db    129  --LIQERREQEDTG------DNQISQQSSSKDHECPEREDLSIHEEPTNSPTNDGVLPSTS  181

Qy    313  S-IDVSPDDVVGI----LGQKFWRARKAIANQ--QRVFAVQLFELHRL--IKVQ-  358
                :                           |::      ::
Db    182  SGIPASSPADMTASSRNPLRKVRYSLRLPAKAAQRLKRIANFQPGSLRRIGLSALKISMD  241

Qy    359  --KLIAASPD---LLLDEISFLGKVSAKSYPVKKL--LPSEFLVKPPLP----HVVVK  405
                           |:::::               :: :
Db    242  CLRLADGSPPQIPVALEDHINIPSTSQENPRDIRSLSHQLPMIPPQQPPQVPSHMLPS  301

Qy    406  QRGDSEKTDQHKMESSAENVGRLSNQCHH----------QOSNYMPFANNP-PASPAPNGYCF  458
                                  |:                  |:|:
Db    302  STGQPGHPSHH-MGPLSQOLLPSGPNQGHHSFQVMVKQEPPSQFTPQPHPMQQTPQQQVL  360

Qy    459  PPQPPSGNHQQWLIPVMSPSEGLIYKPHPG-MAHTGHYGGYYGHYMPTPMVMPQYHPGM  517
                |                           ||
Db    361  PQYPPGMQPHQMHQMRQMTAEEYAQMRAREGFMAAQIKQEVPSSGQPTPV-----PGT  414

Qy    518  GFP--PPGNGYFPPYGMM--PTI------MNPYCSSQQQQQ  548
                                              ||:
Db    415  POPQQITPQPGSLGPMGSLGPPTAPPGSQPMNPQQQRIQQQQAAPSASNSPLLVNLLSN  474

Qy    549  QQPNEQMNQFGHPGN-----------------------LQNTQQQQQRSDNEPAPQQ  582
                |||
Db    475  QQPPQQ--QYMYPGPSAQGLSMQQIAAIQQQQHQQYQQRILQQQQQQAMQQQQQQVQQ  532

Qy    583  QQQ------PTKSYPRARKSRQGSTGSSPSGPQGISGSKSFRPFA  621
                            :||||
Db    533  QQQAPPPTTPNPGHP------QGFPPTNQAGPPGTPG-RPIPPYA  570
```

Search completed: July 29, 2004, 08:24:47
Job time : 88 secs